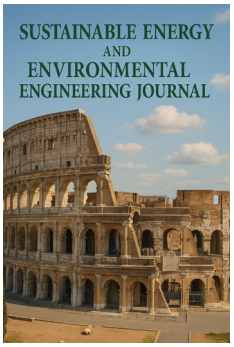
UNIVERSITY OF WEST ATTICA
eISSN:3136-8174

Sustainable Energy and Environmental Engineering Journal

Vol. 4, No. 1, January, 2026



FastSAM DINOv2 for Mine Vegetation Restoration Assessment Using UAV Multispectral and Terrain Data

Konstantin Winter[1,*], Vincent Weber[1], Daniel Buchberger[1]

[1] *Department of Environmental Engineering, Faculty of Civil and Environmental Engineering, TU Wien, Vienna, 1040, Austria*

Received 30 November 2025 Revised 09 March 2026; Accepted 11 March 2026; Published 15 March 2026

Abstract

Monitoring vegetation recovery on reclaimed mine surfaces requires delineating fragmented plant cover under variable substrate, illumination, and terrain conditions. This paper presents a FastSAM DINO framework that combines candidate region generation, frozen DINOv2 appearance descriptors, and independent multispectral and terrain representations. Reliability-gated fusion integrates the three information streams, while dense refinement supports vegetation boundaries and regions omitted by candidate masks. The framework reports vegetation cover, spectral vigor, and structural patch connectivity as separate components before combining them into a restoration screening score. The reported benchmark comprises six mine domains and five spectral bands, with geographically separated evaluation protocols. Full fusion achieves a mean intersection over union of 86.4%, a boundary F1 score of 82.7%, and a mean absolute cover estimation error of 4.8 percentage points. The corresponding appearance-only values are 80.2%, 75.9%, and 7.6 percentage points. The held-mine comparison yields a mean intersection over union of 82.1%, providing a distinct assessment of geographic transfer. Confidence-based deferral relates prediction error within accepted areas to retained valid area and the associated review demand. These comparisons describe mapping performance and the behavior of the proposed screening indicators. The framework connects region segmentation and heterogeneous information fusion to spatially explicit management outputs through observation validity records, spatial partitioning, and uncertainty reporting. Wider ecological interpretation requires independent observations of restoration condition, and operational deployment requires validation of acquisition controls, indicator weights, and processing variability.

Keywords: Mine Rehabilitation, UAV Multispectral Imaging, Terrain Aware Segmentation, Self Supervised Features, Vegetation Recovery Assessment

1. Introduction

Assessment of vegetation restoration distinguishes between the recovery of plant structure and the recovery of overall ecosystem functions on disturbed mine surfaces [1]. Therefore, the monitoring system has added vegetation cover and indicators of establishment status and spatial contiguity [2]. Although the resolution of the UAV observation has improved, interpretation errors may still occur due to inconsistencies in acquisition and space resolution [3]. Terrain-conditioned methods consider the location of slopes and drainage in connecting mapped vegetation with reclamation areas [4]. The aforementioned problems are more severe in areas with narrow benches for young trees, isolated soil pockets or broken strips of exposed spoil. The same vegetation fraction may represent either a continuous patch or several isolated clusters, with different implications for inspection priorities. The assessment product should therefore preserve the spatial arrangement and extent of vegetation recovery. Repeated UAV surveys provide observations of changes within the management areas. Field surveys provide complementary information on species composition and

* Corresponding author: winter.ks@bau.tuwien.ac.at

https://doi.org/10.68406/seeej.2026.vol4iss1nz6:68-82

➢ This is an open access article under the CC BY-NC-SA 4.0 license (https://creativecommons.org/licenses/by/4.0/).

© Authors retain all copyrights.

substrate constraints. Explicit documentation of spatial support links each mapped indicator to the corresponding field observations. Cover estimates quantify the proportion of observable ground occupied by vegetation, while patch metrics describe its spatial arrangement. Spectral measurements add information on the surface condition in the acquisition environment. The assessment must jointly address observation quality, delineation accuracy, and the spatial scope of management decisions. An accountable assessment process links the selected measurements to the spatial units and indicators used in reclamation management.

Multispectral imagery can distinguish surface materials with similar visible appearance. Red-edge and near-infrared measurements are widely used to assess vegetation condition, while illumination, canopy structure, and soil background also influence the measured reflectance. Research on spectral responses in sparse vegetation has shown that the exposed substrate can alter the vegetation signal [5]. Mixed regions containing bright spoil and sparse shrubs introduce classification ambiguity even at fine spatial resolution. Dense shrubs, shadows, and wet ground introduce further variation in surface appearance. Ma et al. [6] addressed the interaction between surface geometry and radiometric variation through a topographic correction framework for high-resolution optical satellite data. Elevation features may encode both site identity and recurring terrain relationships, introducing spatial dependence into model inputs. Spatial dependence also complicates model validation. Adjacent image tiles can share vegetation, shadows, and acquisition artifacts despite having distinct file identifiers. Research on spatial cross-validation has shown that geographically separated evaluation units are required for claims of transfer [7]. The method therefore retains auxiliary information alongside observation quality records and explicit spatial partitioning. Investigations of object-based vegetation mapping have also explored whether coherent regions can improve the interpretation of fragmented cover [8]. A further challenge is to integrate these regions with pretrained visual representations while retaining the identity and reliability of spectral and terrain measurements.

Applying reusable foundation model representations to mine restoration requires explicit interfaces for sensor measurements and assessment outputs. A general region proposal model generates candidate objects whose interpretation as vegetation, rock, or shadow requires task-specific classification. A self-supervised image encoder generates appearance descriptors for subsequent classification and probability calibration. Ecological interpretation also requires independent observations of recovery processes. The adaptation challenge therefore concerns both semantic interpretation and measurement support. Regions that appear contiguous in visible imagery may contain several spectrally distinct materials, while a terrain cell may exceed the size of the vegetation patch of interest. Feature concatenation requires explicit alignment of the spatial support of the input measurements. A research gap concerns the preservation of source-specific measurement identity and quality when integrating information for vegetation assessment across management areas. A related gap is the effect of segmentation error on management indicators. Average overlap can increase even when adjacent patches merge or narrow corridors disappear, altering the spatial organization of recovery. Vegetation classification, boundary agreement, polygon cover, and patch connectivity require separate evaluation to establish their relationships. Geographically independent evaluation is required because familiar surface conditions can inflate estimates of prediction transferability. The evaluation framework should examine how spatial partitioning and indicator construction affect performance and uncertainty. Explicit spatial support, quality control, and interpretation rules link regional vegetation predictions to management indicators. The methodological contribution lies in the interface linking region proposals, source-specific measurement quality, and polygon-level assessment. FastSAM and DINOv2 provide existing visual components, while reliability-aware sensor integration and explicit spatial support define the proposed adaptation to mine restoration.

The assessment problem links the reliability of UAV observations to vegetation delineation and the interpretation of recovery within management units, and the paper is organized around these connected stages. Section 2 examines spectral and terrain-based monitoring together with region segmentation and self-supervised visual representation, identifying the assumptions each family of methods would bring to mine imagery. Section 3 defines the observation units, input conditioning, and candidate region representation. It specifies how visible images remain compatible with pretrained encoders while additional bands and terrain variables enter through separate pathways. Section 4 describes reliability gated fusion, joint supervision, confidence handling, and the construction of restoration indicators. The design preserves the distinction between a vegetation prediction and a management interpretation, so that uncertainty can be inspected before the latter is used. Section 5 presents the experimental specification and examines baseline comparisons, component removal, difficult terrain, geographic transfer, and confidence-based deferral. It also examines the controls used to compare methods across spatial units and input quality conditions. Section 6 states the contribution, limitations, and future research directions. The organization is intended to make the dependencies of every assessment output visible: calibration precedes fusion, independent labels precede evaluation, and ecological interpretation follows independent measurement. Figures are reserved for the input and model designs in the methodological sections and for quantitative displays in the experimental section. This organization establishes the research context before presenting the methodological design and quantitative analysis.

2. Related Work

2.1. Spectral and terrain-based restoration monitoring

Spectral monitoring methods characterize vegetation using measured bands, supervised class predictions, or continuous biophysical proxies. Interpretability derives from the relationship between sensor inputs, acquisition conditions, and surface properties, which supports the diagnosis of classification errors. Research on soil-adjusted spectral indices has examined the impact of bare ground where vegetation cover is low [9]. This sensitivity to soil background is relevant to early reclamation, where young plants occupy a small fraction of the image support. Distinguishing bright spoil, dark coal residue, and wet soil at comparable vegetation fractions requires information beyond a single index threshold. Multiband classifiers can learn more complex relationships when training samples represent substrate and illumination variability. Object-based analysis groups neighboring observations into spatial units for classification. This grouping reduces isolated label noise and makes classification sensitive to the delineation of image regions. Research on object-based environmental classification has investigated the effect of segmentation scale on thematic accuracy [10]. A segmented region may contain shrubs and surrounding substrate with a uniform visible texture, creating ambiguity in plant cover estimation.

Terrain information is commonly introduced through elevation derivatives and neighborhood descriptors. Slope and aspect describe local orientation, while relative elevation and roughness can distinguish positions within an engineered landform. Research on terrain mediated vegetation patterns has investigated how topographic controls interact with environmental conditions and disturbance history [11]. These variables can represent recurring terrain relationships while introducing location specific correlations that reduce geographic generalization. Absolute elevation may identify a particular mine or bench when the intended relationship concerns local exposure or drainage. Terrain surface estimates are subject to errors associated with vegetation, occlusion, and interpolation; vertical accuracy therefore requires independent checkpoint assessment. Monitoring studies consequently need to state whether the terrain layer represents bare ground, canopy surface, or an uncertain mixture. The research gap concerns the joint treatment of spatial support, spectral reliability, and terrain reliability. Interpreting a feature stack requires explicit diagnostics of branch contributions and the treatment of low-quality layers. An assessment method should expose those dependencies and test them under deliberately separated geographic evaluation.

2.2. General segmentation and self-supervised visual representation

General segmentation models separate the generation of regions from the task specific interpretation of those regions. Promptable segmentation research introduced flexible mechanisms for selecting image regions from points, boxes, and other cues [12]. Task specific use requires a defined target class and verification that proposed boundaries correspond to that class. FastSAM formulates general segmentation through a convolutional instance segmentation route and subsequent selection of candidate masks [13]. Its computational motivation makes it relevant to repeated aerial surveys, while multispectral assessment requires additional sensor and semantic interfaces. Narrow plant strips, repeated shrub textures, and patch boundaries formed by mixed soil and foliage can differ substantially from familiar photographic objects. Proposals may overlap, miss thin structures, or group visually similar neighboring surfaces. A downstream method must therefore specify candidate mask filtering, overlap resolution, and the treatment of pixels outside all accepted regions. These rules determine how segmentation choices affect the reported vegetation area. This review emphasizes recent remote sensing adaptations and their relevance to mine imagery; the cited application studies provide context for the adaptation interface.

Self-supervised visual learning provides reusable representations that can reduce dependence on large task specific classification datasets. DINO research examined self-distillation as a way to learn visual features from unlabeled imagery [14]. DINOv2 extended this line of work toward general purpose visual representations learned from curated image collections [15]. These descriptors provide a plausible appearance stream for aerial surface interpretation, with sensitivity to mine substrates, viewing geometry, and seasonal senescence requiring domain specific evaluation. Here, DINO denotes the DINOv2 self-supervised visual encoder, while FastSAM supplies candidate regions. Their distinct roles separate region generation from appearance representation. The proposed combination uses those features as appearance descriptors for subsequent supervised interpretation. Region descriptors are formed from spatial tokens and passed to supervised heads together with independently encoded spectral and terrain inputs. The research gap is the adaptation interface between general image models and calibrated environmental measurements. An effective interface must preserve nonvisual information, recognize poor input quality, and produce outputs that can be evaluated against independent spatial units. Whether the resulting combination improves generalization remains an empirical question requiring actual model training and external observations.

3. Observation Design and Region Representation

3.1. Observation units and input conditioning

The assessment unit is a management polygon with a fixed acquisition date, a known usable area, and a persistent identifier. Prediction operates on image tiles, with cover and connectivity subsequently aggregated to the management polygons. This separation prevents arbitrary tile boundaries from becoming management boundaries. The proposed sensor input contains blue, green, red, red edge, and near infrared reflectance, together with a co-registered terrain grid. Visible reflectance is transformed into a three-channel image for the pretrained encoders using fixed training derived limits. The additional spectral bands enter a dedicated spectral encoder that retains physical channel identity. A calibration record should accompany each flight, including panel observations, exposure information, illumination changes, ground control quality, and the coordinate reference system. Any interpolation or resampling must be documented because it changes the support of both labels and measurements. Spatial support denotes the physical ground area represented by a pixel, token, or region; alignment to a common feature grid preserves a record of each input layer's original support.

Registration checks quantify spatial correspondence before model inference. Control features distributed across the flight block provide horizontal residuals, while independent checkpoints test the terrain surface. Under the quality control protocol, tiles with a horizontal alignment residual above one ground pixel or an invalid reflectance fraction above 5% are flagged for review. The one-pixel alignment criterion and 5% invalid-reflectance cutoff are review settings for this protocol. Their transfer to other sensors requires validation against ground resolution and acquisition conditions. A valid observation mask is retained through every processing stage. Water, deep shadow, and missing terrain retain explicit validity flags within each management polygon. Figure 1 shows the proposed processing order, from flight quality checks and shared spatial partitioning to region generation, feature extraction, and management outputs. The diagram records the proposed design and processing dependencies.

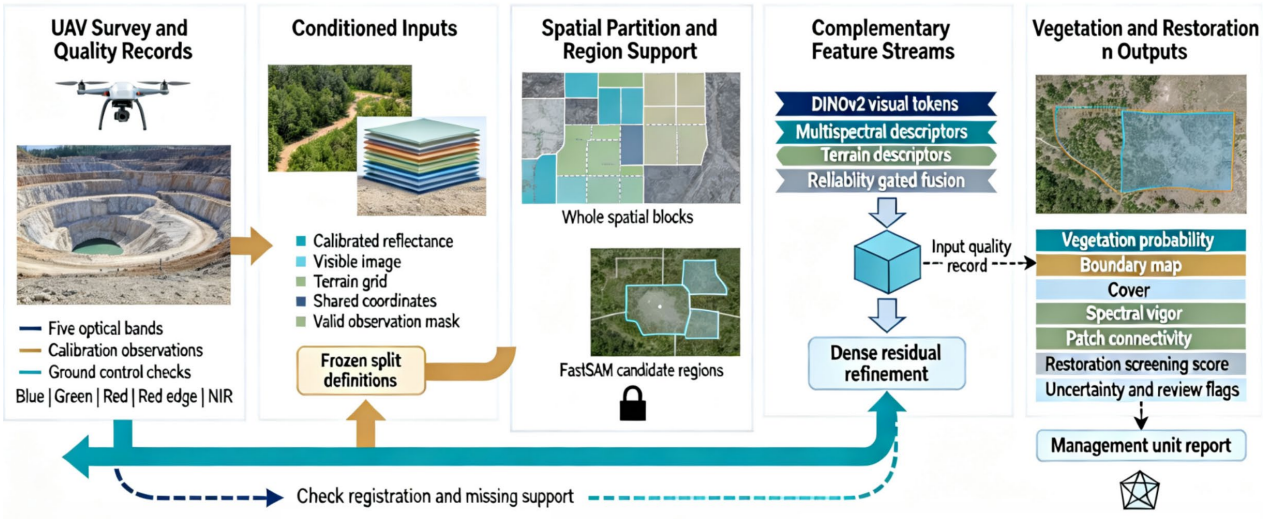


Figure 1. Observation and processing workflow

3.2. Spectral and terrain representations

The spectral branch receives five reflectance channels and two derived contrast channels. One contrast emphasizes the difference between near infrared and red measurements, while the other uses the red edge measurement. The near infrared-red contrast is calculated using the equation below. Reflectance values are dimensionless and remain available alongside the derived channels. The denominator safeguard stabilizes numerical division, while a separate validity mask identifies unreliable radiometry. Here, NIR and R denote near infrared and red reflectance, respectively, and ε is a fixed positive numerical constant.

$$NDVI = \frac{NIR - R}{NIR + R + \varepsilon} \tag{1}$$

Training partition statistics standardize continuous channels, and these statistics are frozen for validation and test partitions. The terrain representation includes slope, sine and cosine of aspect, local roughness, and relative elevation. Aspect encoding avoids a discontinuity between orientations close to north. Relative elevation uses a neighborhood reference at a declared physical radius to reduce reliance on site identity. Roughness is calculated from detrended elevations over the same physical support. Missing terrain values carry validity indicators that distinguish imputed support from measured flat surfaces. The terrain grid is treated as an estimated surface whose correspondence to bare

ground requires checkpoint assessment. Vegetation occlusion and interpolation can affect the derived terrain descriptors and therefore remain part of their quality record. An additional quality vector records local registration residual, reflectance validity, shadow fraction, and terrain support. This vector conditions fusion, while independently annotated vegetation status supplies the prediction target.

Each branch uses a small convolutional encoder to align its output grid with the visual tokens. The alignment procedure retains the original resolution and validity information associated with each input. To examine sensitivity to differences in optical and terrain support, the experimental specification varies terrain resolution and introduces controlled horizontal shifts before feature alignment.

3.3. Candidate regions and visual descriptors

FastSAM supplies candidate masks from the visible image. The model configuration uses its larger proposal variant with a fixed input size of 1024 pixels and retains masks that satisfy validation selected confidence and area limits. These proposed defaults define the starting configuration for validation. Area filtering uses physical units and the minimum mapping target to preserve small regions containing sparse recovery. Duplicate proposals are removed using overlap and confidence, while partially overlapping vegetation candidates remain available until pixel arbitration. The frozen DINOv2 small vision transformer, with a patch size of fourteen, provides dense visual tokens from a separately resized visible image. Token coordinates are mapped to the shared feature grid before pooling. Model names, checkpoint identifiers, resizing rules, and software versions must be fixed during implementation.

For each candidate region, a mask weighted pooling equation aggregates the aligned visual tokens into a regional descriptor. Let M_{jp} represent the soft membership of grid position p in region j, and let $\mathbf{F}_p$ denote the visual feature vector at that position. The resulting descriptor is $\mathbf{d}_j$. A support threshold removes effectively empty masks before aggregation, while ε provides numerical stability. Soft boundary weights distribute coarse token support across neighboring patches.

$$\boldsymbol{d}_j = \frac{\sum_p M_{jp} \boldsymbol{F}_p}{\sum_p M_{jp} + \varepsilon} \tag{2}$$

Equivalent pooling produces spectral and terrain descriptors. A region quality record retains area, perimeter, proposal confidence, and the fraction of valid support for each branch. Proposal confidence describes mask generation quality and supports proposal screening and overlap arbitration. The supervised fusion head produces vegetation probability, which supplies vegetation classification and cover estimation. A separate dense residual head predicts vegetation status for thin structures and pixels outside the candidate regions. Overlap arbitration averages compatible region logits using confidence and valid support, then combines them with the residual prediction. Invalid pixels remain outside the denominator of cover calculations. Boundaries are mapped back to the original grid through an explicitly recorded inverse transform, preserving the relationship between regional decisions and management polygon measurements.

4. Reliability Gated Fusion and Restoration Assessment

4.1. Fusion architecture and decision paths

The fusion module projects visual, spectral, and terrain descriptors into a common latent dimension. A gate receives the region descriptors and their quality record, then produces one reliability weight for each input stream. The weights are normalized across available streams, with unavailable streams excluded before normalization. A missing stream receives zero contribution, and the available-stream weights sum to one. Regions with no valid input support retain an invalid flag and are excluded from valid-area aggregation. This treatment distinguishes missing terrain from reliable evidence of a flat surface. The fusion equation combines the projected descriptors according to their regional reliability weights. Here, $\mathbf{h}_j$ denotes the fused representation, α_{jb} the gate weight, $\mathbf{W}_b$ a learned projection, and $\mathbf{d}_{jb}$ the descriptor from branch b. The branch labels v, s and t identify visual, spectral, and terrain information, respectively.

$$\boldsymbol{h}_j = \sum_{b \in \{v,s,t\}} \alpha_{jb} \boldsymbol{W}_b \boldsymbol{d}_{jb} \tag{3}$$

The gate should be evaluated for sensitivity to degradation in input quality. During training, controlled branch dropout and bounded registration perturbations discourage dependence on a fixed input combination. These perturbations are applied to training units. Evaluation includes controlled input removal and a declared quality perturbation schedule, allowing the response of the fusion mechanism to be examined directly. The resulting diagnostics should relate changes in gate weights to changes in prediction quality under the same observation conditions.

Figure 2 details the three encoders, region pooling, reliability gate, vegetation and boundary heads, dense residual path, and uncertainty aware aggregation. The visible input feeds both proposal generation and the frozen visual encoder,

while calibrated reflectance and terrain retain separate inputs. A training implementation would use a latent dimension of 128, two small hidden layers in the gate, and separate vegetation and boundary heads. Region vegetation predictions and dense residual predictions share the same binary target, with overlap arbitration performed before polygon aggregation. The dense pathway provides additional spatial support for narrow vegetation structures and regions omitted during proposal generation.

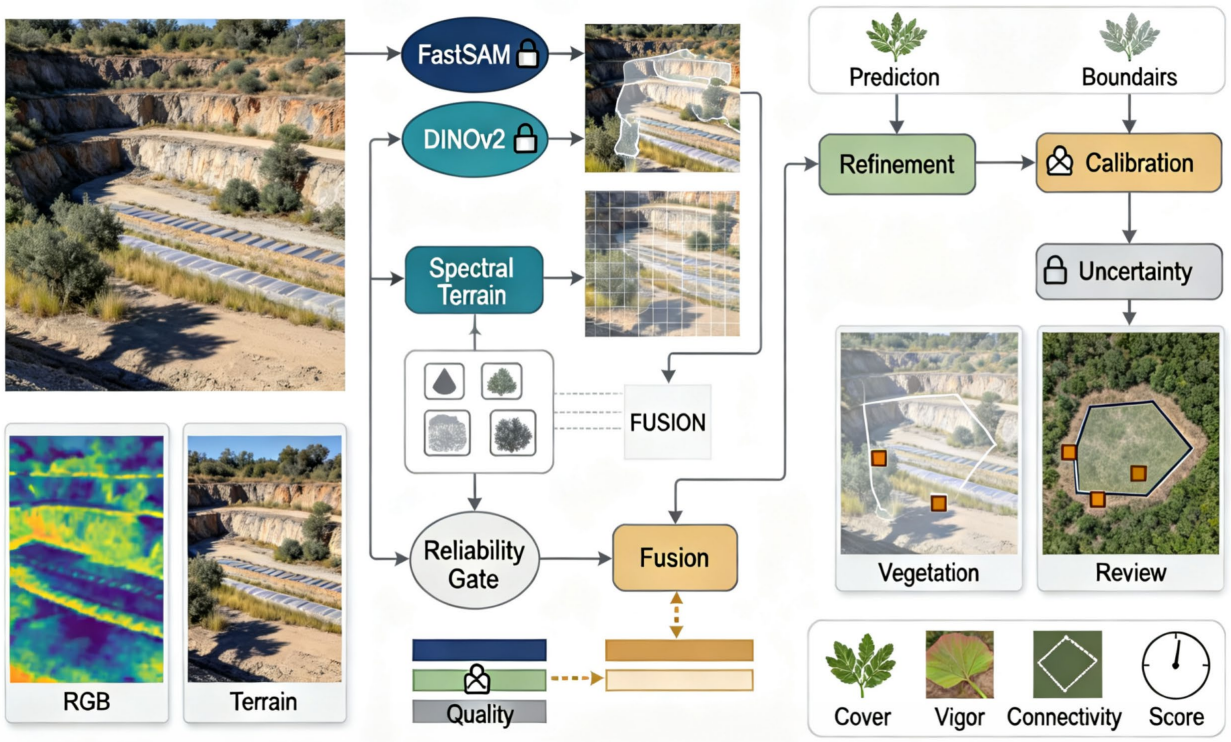


Figure 2. Proposed FastSAM DINO architecture

4.2. Supervision and confidence handling

The proposed training objective combines segmentation, boundary, and consistency losses. Segmentation uses a class weighted pixel classification term and a soft overlap term. Boundary supervision is calculated within a tolerance specified in ground units. Consistency encourages stable predictions under modest radiometric changes and branch dropout when the observation remains valid. Its supervision uses observations from the same acquisition date to preserve genuine temporal change. Separate coefficients enter the training equation so that the contribution and selection of each loss remain explicit. In this notation, L denotes the total loss, λ represents a nonnegative coefficient, and the subscripts identify segmentation, boundary, and consistency components.

$$L = \lambda_{seg} L_{seg} + \lambda_{bd} L_{bd} + \lambda_{con} L_{con} \tag{4}$$

The initial loss weights are 1.0,0.3, and 0.1, respectively. Validation units determine whether these settings require adjustment, while test units remain independent of that selection. Class weights are computed from training labels and capped to limit unstable gradients associated with rare categories. The frozen visual backbone and proposal model are excluded from gradient updates. The branch encoders, projection layers, gate, and prediction heads are optimized jointly. A proposed schedule uses an adaptive optimizer with an initial learning rate of 0.0001, weight decay of 0.01, and a maximum of eighty epochs. Early stopping follows the declared validation overlap objective, while boundary agreement remains a separate diagnostic.

Predicted probabilities require calibration before being interpreted as confidence. A single temperature is fitted on held out validation logits after architecture selection. For a calibrated vegetation probability q, the uncertainty equation takes the form of normalized binary entropy. The uncertainty measure U approaches its limiting values at zero and one by continuity. Polygon uncertainty is calculated from area weighted pixel entropy and reported alongside a registration quality flag. High uncertainty triggers a review flag while preserving the vegetation estimate.

$$U = -\frac{q \ln q + (1-q) \ln(1-q)}{\ln 2} \tag{5}$$

Confidence based deferral changes the evaluated population. Reporting therefore pairs accepted prediction error with the fraction of retained area and separately records performance over all valid pixels. Calibration is assessed through

reliability diagrams and probability bins with support counts. Selective risk is evaluated across a prespecified threshold grid, allowing the error reduction associated with deferral to be considered together with the resulting review workload.

4.3. Restoration indicators and aggregation rules

The assessment report separates vegetation extent from the interpretation of restoration condition. Cover is the predicted vegetation area divided by valid management area, using soft probabilities for estimation and a declared threshold for the binary map. Vigor is a bounded spectral indicator summarized over predicted vegetation, with substrate specific limits estimated from training observations. Connectivity represents the share of vegetation belonging to connected components above a declared minimum area. Its neighborhood definition remains constant across acquisition dates. This component describes mapped patch structure, while ecological connectivity requires independent field interpretation. Spectral vigor similarly characterizes image-derived vegetation condition within the acquisition setting. The restoration score follows a weighted aggregation equation in which cover C, vigor V, and connectivity K are normalized to the interval from zero to one. The resulting screening score R ranges from zero to one hundred.

$$R = 100(0.5C + 0.3V + 0.2K) \tag{6}$$

The weights place the largest emphasis on cover because it is the most directly observable component, while retaining sensitivity to isolated or spectrally weak recovery. Their interpretation reflects the monitoring objective and requires agreement with the intended management use. The weights of 0.5, 0.3, and 0.2 express the declared monitoring preference for cover, vigor, and connectivity, respectively. A sensitivity analysis should vary these preferences within a declared range and identify polygons whose ranking changes substantially. Native species recovery, soil development, long term persistence, and invasive plant occurrence require independent observations alongside the screening score. The score prioritizes inspections and organizes repeated mapping products within a defined restoration assessment procedure.

Aggregation records the amount of valid area, the fraction deferred for review, and the contribution of each component. Polygons with insufficient valid coverage carry a completeness flag during ranking. Change assessment requires comparable seasonal windows and a consistent intersection of valid observation areas. Differences in shadow coverage or exclusion masks can affect the apparent trajectory of recovery. Changes in excluded area should therefore be reported as coverage changes before score differences are interpreted ecologically. Component maps are preserved with the final score so that a change can be examined in terms of additional vegetation cover, altered spectral response, or a different spatial arrangement of patches.

5. Experimental Specification and Quantitative Analysis

5.1. Benchmark design and evaluation controls

The experimental specification defines the observation units, spatial partitions, and evaluation criteria used to organize the analysis. Table 1 lists six mines, A through F, with usable areas of 24, 28, 31, 26, 33, and 30 hectares. Fifty spatial blocks per mine provide three hundred blocks and 1,800 image tiles across the six domains. Optical and terrain grids are specified at five and twenty centimeters, respectively, before alignment. UAV radiometric assessment protocols treat calibration records and illumination conditions as part of the reflectance measurement chain [16]. The proposed acquisition record follows that principle by retaining panel observations, exposure settings, and checkpoint residuals with each input layer. Within each mine, whole blocks are allocated 60% to training, 20% to validation, and 20% to testing, with ten-meter exclusion buffers between partitions. A separate geographic protocol rotates a held test mine, a validation mine, and four training mines through six folds. The within-mine allocation corresponds to thirty training, ten validation, and ten test blocks per mine. The geographic protocol assigns entire mines to these roles and is reported separately from within-mine partitioning. Stratified reference sampling addresses the representation of rare or difficult mapping conditions [17]. Vegetation density, slope, and shadow fraction consequently define the proposed annotation strata, while the block identifiers preserve their geographic context.

Table 1. Mine domain inventory and observation design

Mine	Area (ha)	Blocks	Tiles	Shadow (%)
A	24	50	300	8
B	28	50	300	12
C	31	50	300	17
D	26	50	300	21
E	33	50	300	26

F	30	50	300	31
Total	172	300	1800	Not pooled

Spatially dependent map evaluations require resampling units that reflect the observation dependence structure [18]. Each block therefore remains intact during partitioning and any future uncertainty estimation; its six image tiles share a common block identifier. The specified metrics comprise mean intersection over union across vegetation and nonvegetation, vegetation F1, boundary F1 within a ten-centimeter tolerance, and polygon cover mean absolute error in percentage points. A percentage-point difference is an absolute difference between percentage-valued quantities; a relative percentage reduction divides the absolute change by its stated baseline. Probability calibration assessment distinguishes confidence reliability from class discrimination [19]. Ten equal width probability bins, together with their support counts, provide the corresponding diagnostic. Table 2 lists the baseline performance aggregates and timing budgets used throughout the analysis. The figure set presents aggregate scores alongside distributions of vegetation contrast, polygon cover error, and spatial variation. A five-seed evaluation protocol specifies how sensitivity to initialization should be examined. Statistical inference requires seed-wise dispersion and block-resampled uncertainty in addition to aggregate scores, with spatial blocks retained as intact resampling units. The evaluation specification distinguishes aggregate performance, distributional variation, and uncertainty diagnostics. These complementary views support the joint interpretation of aggregate performance and distributional variation.

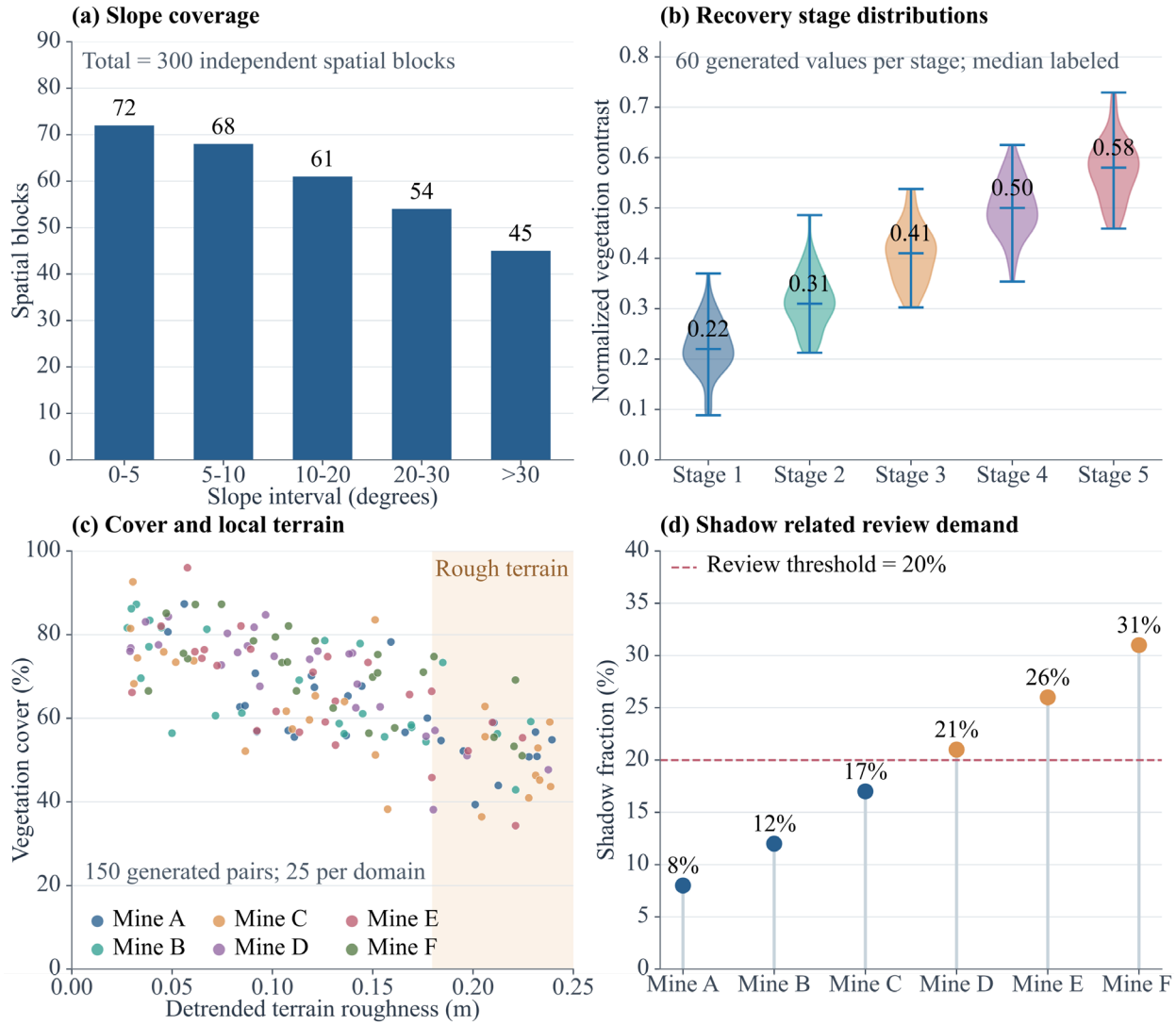


Figure 3. Benchmark characteristics. (a) Slope interval counts. (b) Vegetation contrast distributions across recovery stages. (c) Coverage and terrain roughness classified by mining area. (d) Mine shadow fractions and a quality review threshold

Table 2. Baseline performance and processing budgets

Model recipe	mIoU (%)	Boundary F1 (%)	Cover MAE (pp)	Latency (ms)
Spectral RF	71.8	61.4	10.4	18

Visible U Net	77.5	71.8	8.2	42
Adapted FastSAM	78.6	74.6	7.9	29
DINOv2 dense	79.4	72.3	7.7	61
Appearance fusion	80.2	75.9	7.6	68
Full fusion	86.4	82.7	4.8	79

Figure 3(a) presents 72, 68, 61, 54, and 45 blocks across five slope intervals, placing 99 of the three hundred blocks on slopes above twenty degrees. This distribution gives difficult terrain a substantial evaluation presence while preserving greater support on gentler surfaces.

Figure 3(b) presents vegetation contrast medians of 0.22, 0.31, 0.41, 0.50, and 0.58 across five recovery stages. Their overlapping distributions illustrate why a single threshold may produce stage dependent omissions even when the medians are ordered. Figure 3(c) retains one hundred fifty cover and roughness pairs across the six domains. The visible spread at comparable roughness suggests that cover requires spectral and regional context alongside terrain descriptors. Figure 3(d) presents shadow fractions of 8%, 12%, 17%, 21%, 26%, and 31%; the 20% review threshold flags three domains. Shadow aware mapping methods distinguish observation validity from thematic surface class [20]. Applying that distinction here makes the review flag an observation quality attribute. A stratum with high shadow coverage would require additional reference checks before its vegetation estimates could be interpreted at the same confidence as those from well illuminated ground.

The baseline protocol separates input availability, architecture, and computational demand. It specifies a spectral random forest, a visible image U Net, an adapted FastSAM region classifier, a frozen DINOv2 dense classifier, appearance fusion, and full fusion. Each recipe has declared inputs and uses the same geographic units wherever its learning procedure permits. Reproducible environmental model comparisons document preprocessing, selection rules, and computational conditions alongside performance measures [21]. The proposed implementation record includes normalization statistics, checkpoint identifiers, tuning budgets, and the validation objective for each recipe. A conventional multispectral terrain network and a similarly sized late fusion network would provide additional controls for the benefit of added sensor information. Timing would include loading, proposal generation, feature extraction, arbitration, and polygon aggregation under a declared hardware configuration. Reference boundaries near narrow strips would receive separate review because their small spatial support makes them sensitive to registration tolerance. Polygon size would also remain available during error analysis, allowing percentage errors to be related to the area affected by an inspection decision. Reference sampling should distinguish label uncertainty from input quality. At partially vegetated boundaries, an independent reviewer would record whether disagreement concerns class identity, boundary position, or missing image support. Those categories help determine whether an apparent model error requires a corrected annotation, improved registration, or additional training examples. A spatial audit would also compare the distance of each test block to the nearest training boundary and retain the buffer geometry with the partition file. Evaluation denominators would be checked against the usable area recorded for each mine, making any area excluded through quality screening visible in the reported sample composition.

5.2. Comparative segmentation and indicator behavior

Boundary sensitive segmentation metrics describe contour agreement separately from region overlap [22]. The paired metrics in Figure 4(a) reflect these different error modes: mean intersection over union increases from 71.8% for the spectral random forest to 86.4% for full fusion, while boundary F1 increases from 61.4% to 82.7%. Relative to appearance fusion, full fusion gains 6.2 overlap points and 6.8 boundary points. The contour improvement is compatible with the proposed roles of auxiliary sensing and boundary supervision. Attribution to individual mechanisms requires matched intervention controls that account for their joint contributions and training variability. Figure 4(b) displays signed polygon error distributions for the six recipes, with zero and five percentage point guides. Table 2 lists mean absolute cover errors of 10.4, 8.2, 7.9, 7.7, 7.6, and 4.8 percentage points. The decrease from appearance fusion corresponds to a 36.8% relative reduction in mean absolute cover error. The boxplots describe differences in dispersion and median bias across model recipes, complementing the aggregate errors listed in the table. Individual polygon errors remain relevant even when positive and negative discrepancies produce a small overall bias.

Figure 5(a) presents overlap values of 80.2%, 83.1%, 84.2%, 85.0%, 85.6%, and 86.4% for appearance fusion, removal of terrain, removal of the spectral branch, removal of gating, removal of boundary loss, and full fusion. Relative to the complete design, removing terrain produces the largest single component decrease at 3.3 points, followed by spectral removal at 2.2 points. Gating and boundary supervision correspond to smaller decreases of 1.4 and 0.8 points. These contrasts are absolute score differences between reported retrained configurations. Their interpretation is conditional on the retained components, while branch interactions and training variability require matched factorial controls and repeated runs. Figure 5(b) places those differences within three slope strata. The overlap advantage of full

fusion relative to appearance fusion increases from 4.2 percentage points on low slopes to 6.2 on moderate slopes and 10.7 on high slopes. On the high slope stratum, full fusion exceeds the terrain removal and gate removal settings by 7.1 and 4.2 points. That pattern identifies difficult geometry as a priority for matched testing of terrain support, illumination, and alignment. Direct perturbation tests would assess whether the gate reduces the weights assigned to degraded input streams.

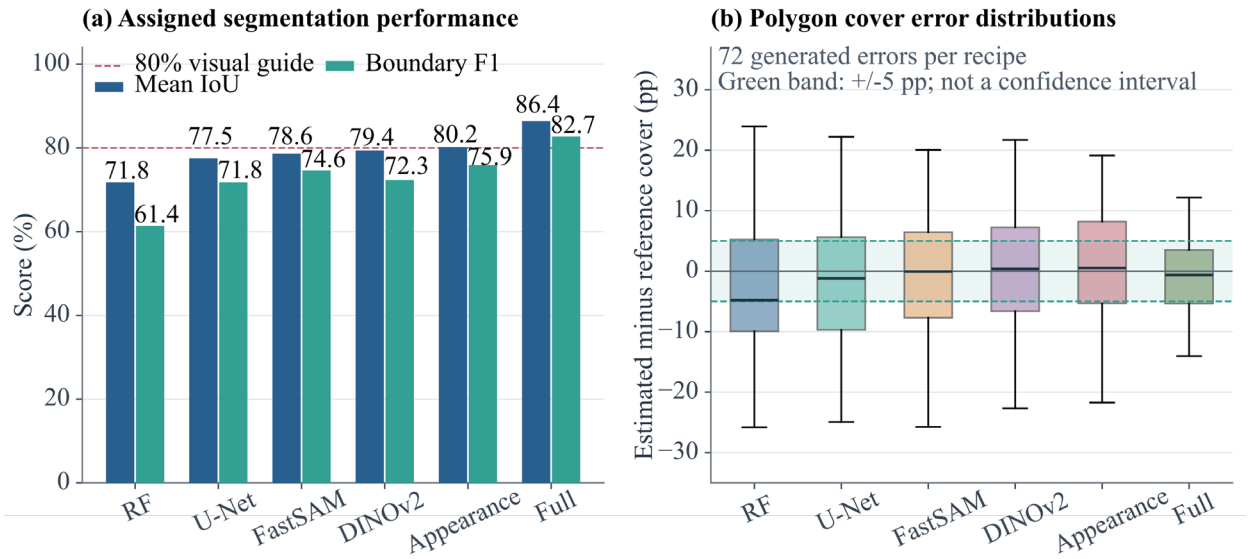


Figure 4. Baseline comparisons and polygon cover error distributions. (a) Mean intersection over union and boundary F1 across six model recipes. (b) Signed polygon cover error distributions with zero and tolerance guides

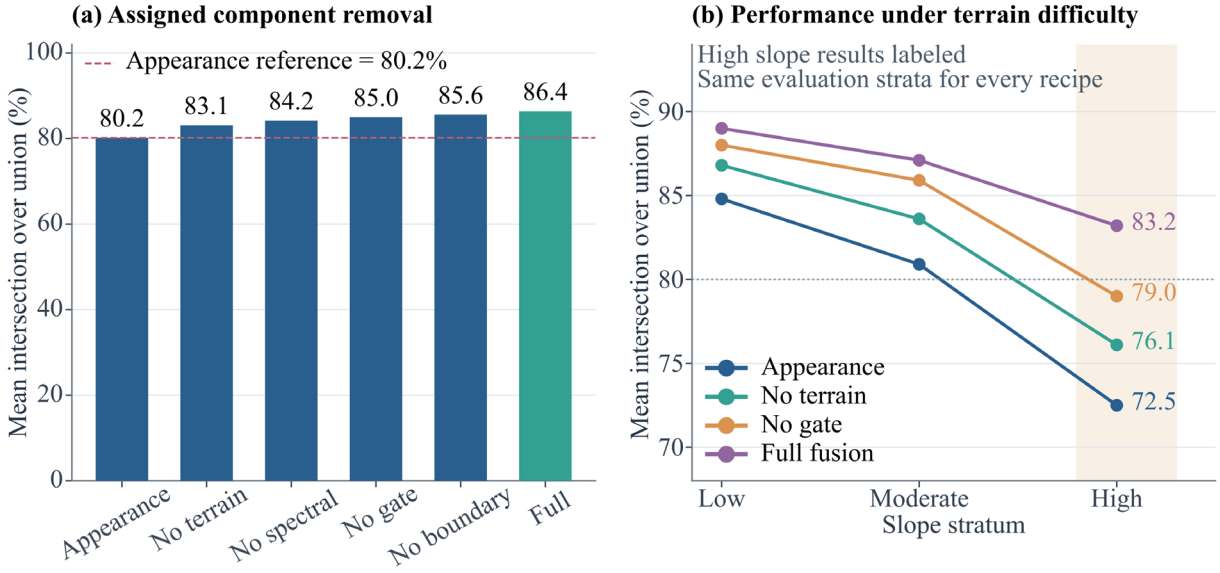


Figure 5. Component removal and terrain sensitivity analysis. (a) Overlap for six component configurations. (b) Overlap across three slope strata for four configurations

Figure 6(a) relates mapping errors to eighty cover estimation pairs spanning approximately 5% to 95% reference cover. Its sample mean absolute error is 4.70 percentage points after predictions are clipped to the valid percentage range. This value summarizes the eighty plotted pairs, whereas the table-level aggregate cover error is 4.8 percentage points. Their respective evaluation supports should accompany the two summaries when interpreting this difference. The small positive sample bias of 0.30 points coexists with individual departures beyond the shaded tolerance band, illustrating the distinction between systematic displacement and typical error magnitude. Figure 6(b) decomposes the restoration score into cover, vigor, and connectivity across six polygons. Connectivity metrics depend on the component definition and spatial support used to construct a landscape graph or patch map [23]. The proposed reporting therefore retains the mapping scale and neighborhood rule with the score. Polygon P3 combines cover of 54, vigor of 61, and connectivity of 49 to produce 55.1; P4 combines 63, 58, and 64 to produce 61.7. Their 6.6-point difference contains contributions of 4.5 from cover, minus 0.9 from vigor, and 3.0 from connectivity. The ordering consequently reflects the weighting of two improving components against one declining component. Component visibility allows a manager to identify the source of the rank difference and assess whether it matches the inspection objective.

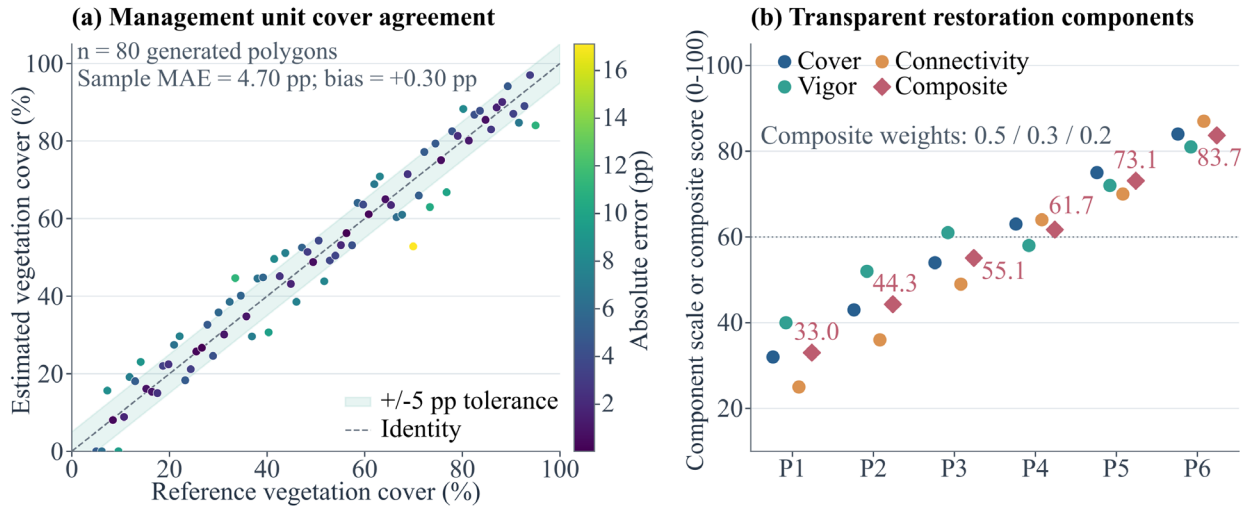


Figure 6. Management indicator analysis. (a) Reference and estimated vegetation cover with an identity line and tolerance band. (b) Cover, vigor, connectivity, and calculated restoration score across six polygons

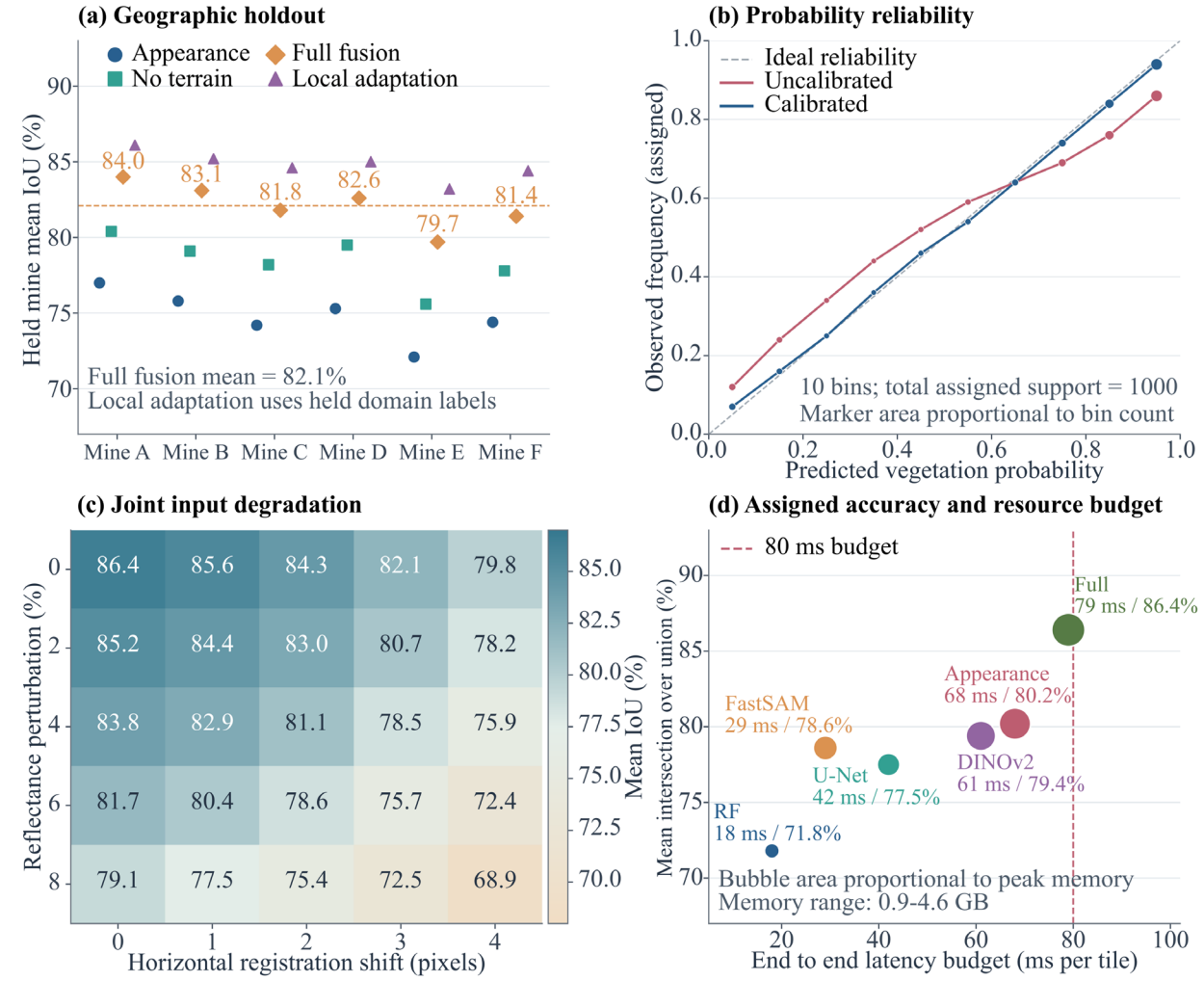


Figure 7. Geographic transfer and robustness diagnostics. (a) Held mine overlap for four adaptation conditions. (b) Probability reliability curves with bin support. (c) Reflectance perturbation and horizontal registration sensitivity surface. (d) End to end latency and overlap tradeoff for six configurations

Error attribution requires examining the spatial locations behind the aggregate metrics. Thin vegetation strips should be evaluated before and after dense refinement to distinguish proposal omissions from classification errors. Overlapping shrub masks should likewise retain their intermediate support so that a connectivity increase can be traced to a genuine corridor or an arbitration artifact. A controlled implementation would pair each ablation across the same polygons and report errors separately for sparse, shaded, and steep strata. Threshold sensitivity should be checked near partially

covered boundaries, where a small probability change may alter both area and patch membership. Polygon reports should retain the usable area denominator and the distribution of local errors, enabling inspectors to identify concentrated discrepancies along drains or bench edges. Independent observations of species composition, ground stability, and establishment persistence would supply the ecological context for those spatially explicit screening outputs. Probability threshold variation provides a further test of indicator stability. A polygon dominated by intermediate probabilities may show substantial cover variation across nearby thresholds, while a polygon with confident vegetation interiors may retain a stable area estimate. Connectivity can react more abruptly when a narrow link changes class, even if the affected area is small. Inspecting those links would distinguish a robust patch relationship from a threshold dependent connection. Paired cover and connectivity sensitivity should therefore accompany score changes near an inspection boundary. Each disputed corridor would retain its width, neighboring probabilities, and local registration quality for subsequent review.

5.3. Transfer robustness uncertainty and deployment constraints

Luo et al. [24] investigated geographic domain shift in remote sensing semantic segmentation through vision foundation model fine-tuning. Figure 7(a) presents six held mine overlap values of 84.0%, 83.1%, 81.8%, 82.6%, 79.7%, and 81.4% for full fusion. Their mean of 82.1% is 4.3 points below the within domain comparison and 7.3 points above the appearance transfer mean. The 4.3 point spread between mines A and E indicates heterogeneous transfer difficulty. Mine E also has the lowest appearance transfer value, making domain specific substrate and acquisition conditions relevant targets for inspection. Local adaptation uses held domain labels and is presented as a separate diagnostic condition. The within-domain and held-mine scores arise from different partitioning protocols, so their difference characterizes transfer difficulty across those protocols. Adapted conditions are distinguished from evaluation conducted without labels from the held test mine.

Figure 7(b) examines probability reliability through ten bins representing one thousand observations. The calibrated curve remains closer to the identity guide; in the highest bin, the observed frequency is 0.94 after calibration compared with 0.86 before calibration, against a predicted probability of 0.95. This visual comparison supports a qualitative assessment of reliability; numerical calibration error and bin-wise uncertainty require the corresponding sample counts. Marker size exposes uneven support across bins. Fitting the temperature on a separate validation mine preserves the geographic separation between calibration and test evaluation.

Registration error analysis treats spatial correspondence as a distinct source of uncertainty in multisensor mapping [25]. Figure 7(c) presents overlap across five reflectance perturbation levels and five horizontal shifts. Increasing shift from zero to four pixels at the reference reflectance level reduces overlap by 6.6 points, while increasing reflectance perturbation from zero to 8% at zero shift reduces it by 7.3 points. The combined extreme produces a 17.5-point decrease, exceeding the sum of the two separate decreases by 3.6 points. This 3.6-point excess describes departure from additive score changes on the tested perturbation grid; estimating a statistical interaction requires repeated perturbation measurements. The joint degradation therefore warrants analysis of radiometric and geometric disturbances together. The reference mask remains fixed during perturbation, preserving an independent spatial target. Figure 7(d) places six configurations against latency budgets of 18, 42, 29, 61, 68, and 79 milliseconds per tile. Relative to appearance fusion, full fusion increases latency by eleven milliseconds and peak memory by 0.5 GB, alongside a 6.2 percentage point increase in overlap. The full fusion latency provides a one millisecond margin relative to the proposed eighty millisecond budget, indicating sensitivity to loading variability and high proposal density. These tile-level values define the reported processing budget. Operational timing additionally requires hardware identifiers, loading and aggregation costs, and upper latency quantiles alongside mean latency.

Figure 8 presents accepted error rates of 9.6%, 8.3%, 7.1%, 6.2%, 5.4%, and 4.1% at retained area fractions of 100%, 95%, 90%, 85%, 80%, and 70%. Selective prediction theory evaluates conditional error jointly with retained coverage [26]. At 85% retention, accepted error decreases by 3.4 percentage points while 15% of valid area is deferred. Reducing retained area from 85% to 80% increases the deferred area by five percentage points and reduces accepted error by a further 0.8 percentage points. A further reduction in retained area from 80% to 70% increases the deferred area by ten percentage points and reduces accepted error by 1.3 percentage points. These conditional differences suggest declining error reduction per additional area deferred over that interval. Retained coverage denotes accepted valid image area, and accepted error is conditional on that retained area. Each operating point therefore retains its own evaluation denominator. The shaded operating range can be interpreted against inspection capacity and the consequence of a missed vegetation boundary. Reports should preserve the full valid area, accepted area, and review area as distinct quantities. Random audits of accepted predictions would assess confident errors within the retained area.

Resource allocation should account for the spatial concentration of uncertain observations. A small uncertain area near a drainage channel may require a different inspection response from an equally uncertain area on stable ground. Polygon records should therefore retain the locations of quality flags, their acquisition provenance, and the corresponding inspection actions. Seasonal comparisons require consistent observation windows, a common valid-area denominator, and recorded moisture conditions. Deployment trials should record review time, correction type, and

observed substrate or illumination conditions. These records can inform the diagnosis of recurring acquisition problems and identify coverage gaps for subsequent sampling. Revised reference labels should be versioned and retained within their original evaluation partitions. Consistent ground support across surveys is required to control interpolation-induced changes in local descriptors and uncertainty at polygon boundaries. Inspectors should have access to the original image support and processing settings for each disputed area. Computational assessment should also retain the distribution of tile processing times. High proposal density, irregular boundaries, and storage latency may produce processing delays beyond the mean evaluation latency. Deployment tests should record upper latency quantiles and the proportion of tiles exceeding the processing limit, then examine their association with difficult vegetation or terrain strata. Field visits should be scheduled by spatial proximity to reduce travel between uncertain areas. Operating cost should be reported per usable hectare alongside accepted and deferred area so that coverage thresholds can be compared under the same inspection capacity. Work records should distinguish time spent on image processing, human interpretation, and field verification to estimate operating costs at a specified coverage threshold.

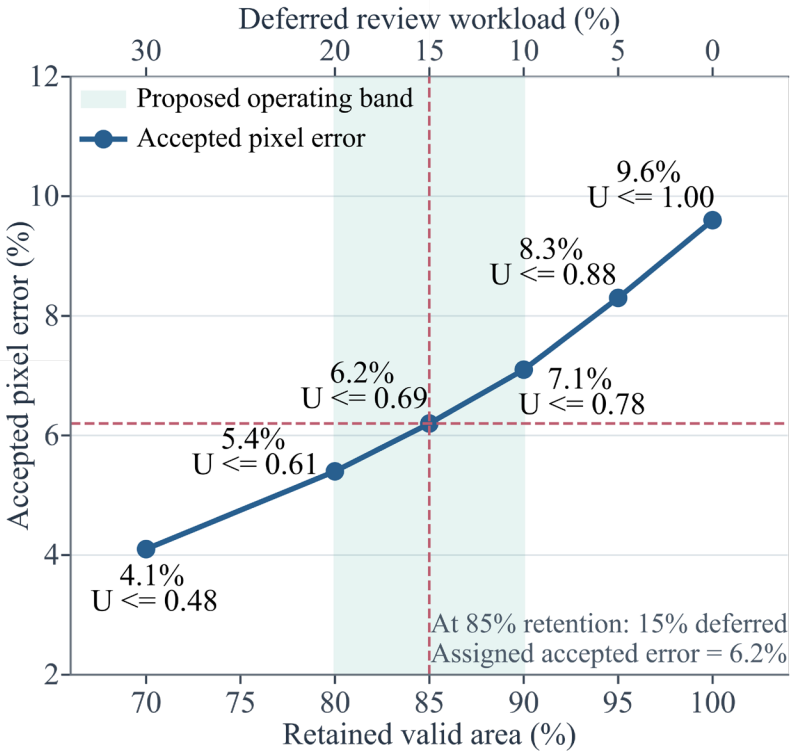


Figure 8. Selective prediction and review workload

6. Conclusion

This paper proposes a FastSAM DINO framework for assessing mine vegetation restoration using UAV multispectral imagery and terrain data. The design separates candidate region generation from semantic interpretation and uses distinct pathways for visible appearance, calibrated reflectance, and terrain context. Reliability gating, dense boundary refinement, and explicit validity handling connect these pathways to vegetation maps and management indicators. Cover, spectral vigor, and patch connectivity are reported separately and then combined into a screening score. The manuscript also defines spatial partitioning, baseline controls, and uncertainty reporting for evaluating the framework, preserving the distinction between mapping quality and ecological recovery.

The main limitations concern the restricted study extent, variability in data acquisition, and interpretation of the composite restoration index. Further evaluation should examine geographic applicability, computational variability, and the correspondence between remotely sensed observations and ecological conditions. Further studies should assess training sensitivity and responses to acquisition errors. Performance on sparse vegetation and irregular spoil surfaces, together with terrain reliability under occlusion, remains an important condition for geographic transfer. Composite index weights should reflect management objectives, with a stated rationale for each application. Ecological interpretation requires observations of vegetation cover and spectral responses across species and ecosystems.

Future work should evaluate the complete pipeline under additional, independently labeled observations from contrasting mines and acquisition periods. Controlled comparisons should supply conventional networks with the same information and tuning effort, allowing architectural effects to be distinguished from additional sensor inputs. Field surveys should evaluate the ecological meaning of indicator components and identify cases requiring supplementary field evidence. Further research should examine uncertainty under unfamiliar substrates, efficient updating after new

surveys, and the consequences of scale for patch connectivity. Transparent configuration records and accessible spatial partitions would support replication and enable operational users to inspect why a particular area was assigned for review.

7. Declarations

7.1 Author Contributions

Konstantin Winter contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision, project administration, and funding acquisition. Vincent Weber contributes to validation, analysis, investigation, data collection, draft preparation. Daniel Buchberger contributes to conceptualization, validation, analysis. All authors have read and agreed with the manuscript before its submission and publication.

7.2 Funding and Acknowledgments

This research received no specific financial support from any funding agency.

7.3 Institutional Review Board Statement

Not applicable.

7.4 Informed Consent Statement

Not applicable.

7.5 Date Availability

Not applicable.

7.6 Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

7.7 Declaration of generative AI and AI-assisted technologies

During the preparation of this manuscript, the authors utilized Grammarly for linguistic polishing, logical organization and draft revision. The generative AI tool was not involved in research design, data collection, data analysis, or the equation of core conclusions. All outputs generated by AI were reviewed, revised and verified manually by the authors. The authors take full responsibility for all content of this manuscript.

8. References

[1] Tibbett, M. (2024). Post-mining ecosystem reconstruction. Current Biology, 34(9), R387-R393. https://doi.org/10.1016/j.cub.2024.03.065

[2] Chen, Z., Hou, H., Zhang, S., Campbell, T., Yang, Y., Tu, M., Yuan, Y., & Dixon, K. W. (2024). Using unmanned aerial vehicle multispectral data for monitoring the outcomes of ecological restoration in mining areas. Land Degradation & Development, 35(4), 1599–1613. https://doi.org/10.1002/ldr.5010

[3] Singh, K. K., Surasinghe, T. D., & Frazier, A. E. (2024). Systematic review and best practices for drone remote sensing of invasive plants. Methods in Ecology and Evolution, 15(6), 998–1015. https://doi.org/10.1111/2041-210x.14330

[4] Tripathi, N., Singh, R. S., Chaulya, S. K., & Hills, C. D. (2025). Recovery of damaged mine spoil: An ecological & geotechnical perspective. Journal of Soils and Sediments, 25(11), 3439–3459. https://doi.org/10.1007/s11368-025-04170-z

[5] Bhadra, S., Sagan, V., Sarkar, S., Braud, M., Mockler, T. C., & Eveland, A. L. (2024). PROSAIL-Net: A transfer learning-based dual stream neural network to estimate leaf chlorophyll and leaf angle of crops from UAV hyperspectral images. ISPRS Journal of Photogrammetry and Remote Sensing, 210, 1–24. https://doi.org/10.1016/j.isprsjprs.2024.02.020

[6] Ma, Y., Liang, S., Peng, W., He, T., Ma, H., Chen, Y., Li, W., Xu, J., & Guan, S. (2025). A universal physically-based topographic correction framework for high-resolution optical satellite data. ISPRS Journal of Photogrammetry and Remote Sensing, 227, 459–480. https://doi.org/10.1016/j.isprsjprs.2025.05.027

[7] Linnenbrink, J., Milà, C., Ludwig, M., & Meyer, H. (2024). kNNDM CV: k-fold nearest-neighbour distance matching cross-validation for map accuracy estimation. Geoscientific Model Development, 17(15), 5897–5912. https://doi.org/10.5194/gmd-17-5897-2024

[8] Gui, B., Sam, L., Bhardwaj, A., Gómez, D. S., Peñaloza, F. G., Buchroithner, M. F., & Green, D. R. (2025). SAGRNet: A novel object-based graph convolutional neural network for diverse vegetation cover classification in remotely-sensed imagery. ISPRS Journal of Photogrammetry and Remote Sensing, 227, 99–124. https://doi.org/10.1016/j.isprsjprs.2025.06.004

[9] Paul, N. C., Ponnaganti, N., Gaikwad, B. B., Sammi Reddy, K., & Nangare, D. D. (2025). Optimized soil adjusted vegetation index mapping of Pune district using Google Earth Engine. Remote Sensing Letters, 16(7), 728–736. https://doi.org/10.1080/2150704x.2025.2502176

[10] Gui, B., Sam, L., Bhardwaj, A., Gómez, D. S., Peñaloza, F. G., Buchroithner, M. F., & Green, D. R. (2025). Optimizing object-based agricultural and ecological land use classification via multi-band fusion and seasonal spectral differences. International Journal of Remote Sensing, 46(16), 6157–6188. https://doi.org/10.1080/01431161.2025.2532837

[11] Lai, J., Li, J., & Liu, L. (2025). Optimization strategies for ecological security pattern based on the remote sensing ecological index in Yunnan Province, China. Land Degradation & Development, 36(4), 1326–1342. https://doi.org/10.1002/ldr.5432

[12] Wang, C., Chen, J., Meng, Y., Deng, Y., Li, K., & Kong, Y. (2024). SAMPolyBuild: Adapting the Segment Anything Model for polygonal building extraction. ISPRS Journal of Photogrammetry and Remote Sensing, 218, 707–720. https://doi.org/10.1016/j.isprsjprs.2024.09.018

[13] Ding, L., Zhu, K., Peng, D., Tang, H., Yang, K., & Bruzzone, L. (2024). Adapting Segment Anything Model for change detection in VHR remote sensing images. IEEE Transactions on Geoscience and Remote Sensing, 62, 1–11. https://doi.org/10.1109/tgrs.2024.3368168

[14] Wang, Y., Albrecht, C. M., & Zhu, X. X. (2024). Multilabel-guided soft contrastive learning for efficient Earth observation pretraining. IEEE Transactions on Geoscience and Remote Sensing, 62, 1–16. https://doi.org/10.1109/tgrs.2024.3466896

[15] Ray, I., & Skurikhin, A. (2025). Deep clustering of remote sensing scenes through heterogeneous transfer learning. International Journal of Remote Sensing, 46(8), 3089–3109. https://doi.org/10.1080/01431161.2025.2465917

[16] Song, S., & Qin, R. (2024). A general albedo recovery approach for aerial photogrammetric images through inverse rendering. ISPRS Journal of Photogrammetry and Remote Sensing, 218, 101–119. https://doi.org/10.1016/j.isprsjprs.2024.09.001

[17] Yi, C., Zhao, X., Sun, Q., & Wang, Z. (2024). Assessing the accuracy of remote sensing data products: A multi-granular spatial sampling method. Future Generation Computer Systems, 159, 151–160. https://doi.org/10.1016/j.future.2024.04.062

[18] Arthur, R. (2025). A general method for resampling autocorrelated spatial data. Geographical Analysis, 57(2), 302–319. https://doi.org/10.1111/gean.12417

[19] Balanya, S. A., Maroñas, J., & Ramos, D. (2024). Adaptive temperature scaling for robust calibration of deep neural networks. Neural Computing and Applications, 36(14), 8073–8095. https://doi.org/10.1007/s00521-024-09505-4

[20] Chi, K., Li, J., Jing, W., Li, Q., & Wang, Q. (2024). Neural implicit Fourier transform for remote sensing shadow removal. IEEE Transactions on Geoscience and Remote Sensing, 62, 1–10. https://doi.org/10.1109/tgrs.2024.3412368

[21] Filatovas, E., Stripinis, L., Orts, F., & Paulavičius, R. (2024). Advancing research reproducibility in machine learning through blockchain technology. Informatica, 35(2), 227–253. https://doi.org/10.15388/24-infor553

[22] Wang, X., Li, Y., & Shi, A. (2025). SAEENet: A semantic adaptive edge enhancement network for remote sensing image semantic segmentation. Remote Sensing Letters, 16(10), 1109–1119. https://doi.org/10.1080/2150704x.2025.2536240

[23] He, J., Li, Y., Zhou, W., Qian, M., Zha, E., & Shi, X. (2025). A multi-objective framework for regional ecological planning: Restoration prioritization analysis. Land Degradation & Development, 36(11), 3670–3684. https://doi.org/10.1002/ldr.5591

[24] Luo, M., Zan, Y., Khoshelham, K., & Ji, S. (2025). Domain generalization for semantic segmentation of remote sensing images via vision foundation model fine-tuning. ISPRS Journal of Photogrammetry and Remote Sensing, 230, 126–146. https://doi.org/10.1016/j.isprsjprs.2025.09.004

[25] Huang, J., Yang, F., & Chai, L. (2024). Multimodal remote sensing image registration based on adaptive spectrum congruency. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 17, 14965–14981. https://doi.org/10.1109/jstars.2024.3411706

[26] Hendrickx, K., Perini, L., Van der Plas, D., Meert, W., & Davis, J. (2024). Machine learning with a reject option: A survey. Machine Learning, 113(5), 3073–3110. https://doi.org/10.1007/s10994-024-06534-x