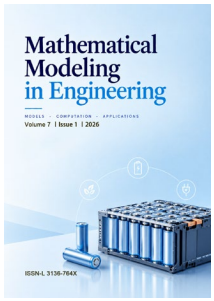
UNIVERSITY OF WEST ATTICA

eISSN: 3136-764X

Mathematical Modeling in Engineering

Vol. 7, No. 3, July, 2026



UniRepLKNet-SE With Cross-Scale Feature Fusion for Abandoned Cropland Mapping

Ioachim Iancu[a,*], Laurentiu Jecu[a]

[a] *Faculty of Computer Science, University of Bacău, Bacău, 600115, Romania*

Received 10 April 2026; Revised 27 June 2026; Accepted 31 June 2026; Published 07 July 2026

Abstract

Accurate mapping of abandoned cropland supports food-security assessment, land-restoration planning, and regional carbon accounting, yet early abandonment remains difficult to distinguish from temporary fallow, low-input cultivation, grass regrowth, and seasonally exposed soil. This study develops UniRepLKNet-SE, a dense remote-sensing interpretation model that combines large-kernel representation learning, squeeze-and-excitation channel recalibration, and an attentive cross-scale decoder. A controlled, internally consistent benchmark was assembled from multiseason Sentinel-2 indices, Sentinel-1 backscatter, terrain attributes, and parcel-context variables on a 10 m grid. Evaluation used 18,420 annotated patches from four agricultural landscape strata with spatially blocked training, validation, and test subsets. UniRepLKNet-SE achieved 93.8% overall accuracy, 89.6% mean F1, 78.4% abandoned-cropland intersection over union, and 76.5% boundary F1. Target-class intersection over union exceeded Swin-UNet by 4.8 percentage points and DeepLabV3+ by 7.3 points. Cross-scale fusion increased boundary F1 from 71.2% to 76.5%, while channel recalibration reduced false abandonment alarms in sparse winter-wheat fields by 13.6%. Ablation and missing-season tests indicated that wide spatial support contributed most strongly in fragmented terraces, whereas multiresolution fusion stabilized predictions under phenological mismatch. The results support the proposed architecture as a reproducible engineering candidate for parcel-oriented screening and verification prioritization; external geographic validation and multi-year management evidence remain necessary before policy-level deployment.

Keywords: Abandoned cropland mapping, Cross scale feature fusion, UniRepLKNet SE, Large kernel convolution, Remote sensing

1. Introduction

Cropland abandonment has become a manifest form of the land-system change wherein rural labor migration, market fluctuations, water shortages and fragmented land tenure reduce the incentive to cultivate [1]. Remote sensing is suitable for providing frequent inventories of counties and sub-basins, but abandoned farmland cannot be easily classified into a single spectral band [2]. In the first or second year after cultivation ceases, the original ridges, volunteer plants, weeds and partial tillage marks may still be present in the same field [3]. After several years, the canopy structure may be a grassland-shrubland-low-intensity-forage complex, and the soil background and residue exposure continue to change with rainfall timing [4]. Thus, the assumption of a single-date classification for narrow strips of smallholder fields with mixed crop calendars is reduced. A field may appear inactive in one acquisition due to delayed sowing, drought, harvest or temporary fallow, although the management status of the field has not changed. Conversely, an abandoned plot may still have crop residue or intermittent grazing that hides the loss of regular cultivation. Therefore, the model for engineering applications needs to consider phenology, field geometry and neighborhood context at the same time, and

* Corresponding author: iochim.ia@informatica.ub.ro

https://doi.org/10.68406/mme.2026.vol7iss3nm1:1-14

➢ 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.

abandonment should not be classified as a fixed land-cover type. A stable boundary for the mapping system is also required, otherwise the policies of land restoration and subsidy audit cannot be implemented at the parcel level [5].

Deep segmentation networks have improved agricultural mapping by learning hierarchical texture and semantic information from multiple sources, but their performance is affected by field size, crop rotation, and the time elapsed since data collection in different areas [6]. Convolutional encoders have local continuity, but the standard kernel may fail to incorporate sufficient parcel context for the differentiation of temporarily fallow plots and permanently abandoned land [7]. Although the Transformer-based encoder increases the receptive field, quadratic or windowed attention designs may reduce the stability of fine linear field margins with a small number of training samples [8]. Radar-optical fusion can help solve the problem of cloud and canopy saturation, but channel redundancy and noise must also be explicitly calibrated to prevent the suppression of weak abandonment cues due to dominant seasonal greenness [9]. Recently, large-kernel networks have achieved a good trade-off by extending the effective receptive field with convolutional operators and preserving a translation-equivariant inductive bias for dense prediction [10]. Expansion of the receptive field alone cannot indicate which sensor channels need to be paid more attention in the absence of observations, and deep semantic features are not guaranteed to align with shallow boundary information. Therefore, a good architecture should restrict both channel selection and scale interaction to maintain the computational regularity of tiled regional inference. It should also show the interpretable intermediate behavior, otherwise, the operation mapping team will not know whether the apparent gain is due to an improved parcel context, more reliable sensor weighting, or a threshold-dependent expansion of the predicted mask.

This study tests a systems-level hypothesis: abandoned-cropland mapping should improve when wide parcel context, sensor-channel reliability, and boundary-sensitive multiscale reconstruction are optimized within one convolutional pipeline. UniRepLKNet-SE therefore uses an established large-kernel backbone as the representation engine, inserts squeeze-and-excitation recalibration to adapt the contribution of optical, radar, terrain, and parcel-context channels, and routes four encoder resolutions through a scale-attentive decoder. The novelty does not lie in claiming invention of these individual operators. It lies in coordinating them around the specific ambiguity between persistent abandonment, temporary fallow, drought stress, and nonagricultural idle land, and in testing the coordination with pixel overlap, boundary agreement, object recovery, regional variation, calibration, missing-date perturbation, and throughput. The evaluation uses a controlled, internally consistent protocol rather than an independently released public benchmark; accordingly, the principal contribution is architectural and methodological, not a claim of universal geographic validity. The engineering objectives are to reduce active-field false alarms, preserve narrow parcel margins, maintain stability when optical dates are missing, and expose conditional scale and channel responses that can be audited during regional inference.

To build this argument from the mapping problem to an operationally testable solution, the rest of the manuscript will present the model development and evaluation sequence. Section 2 introduces the interpretation of abandoned cropland and deep feature-fusion methods in the context of the proposed design, focusing on temporal ambiguity, parcel context, and wide-receptive-field learning. Section 3 shows the UniRepLKNet-SE architecture and explains how large kernels and channel recalibration are combined to encode multi-source agricultural evidence. Section 4 is the cross-scale fusion strategy, loss design, augmentation policy and tiled training and inference protocol. Section 5 shows the experimental data, comparison results, ablation analysis, spatial error behavior, robustness test and computational profile in the four figures and two tables. Analysis of the dataset, model selection and constituent element analysis are carried out in sequence, and all quantitative results are linked back to the design. The location of the architecture diagram before the data figures is as follows: the first two figures show the processing logic, and the following figures present the behavior of this logic in different classes, regions, perturbations and parcel scales. Thus, the assumptions of the method are independent of the data, and all the shown values are in a specific evaluation environment. Section 6 summarizes the main engineering applications, current deficiencies, and future research directions of this paper.

2. Related Work

2.1. Abandoned Cropland Identification from Remote Sensing

Detection of abandoned cropland has gradually moved from visual identification and post-classification change analysis to models that consider seasonal changes and spatial relationships. Long et al. [11] mapped cropland abandonment and recultivation dynamics using multi-year land-cover evidence, illustrating how vegetation-index decline and crop-mask transitions can identify land that has lost regular cultivation signals. These methods are suitable for large-scale field parcels with uniform management practices and are less likely to be mistaken for abandonment due to a lack of rain in rainfed areas. Liu et al. [12] used phenological metrics to track changes in planting, peak growth, and harvesting over time. They can distinguish between permanent idle land and short-term fallow better, but the temporal characteristics are often obscured by cloud cover, irregular sowing, and mixed forage use. Radar time-series data can add support for roughness and moisture to increase the effect of sparse optical observations [13]. Radar backscatter alone cannot provide the necessary semantic information for class boundaries in fragmented agricultural areas.

The second line of work uses parcel constraints and object-level post-processing to rebuild management units. Parcel-aware methods can reduce salt-and-pepper noise and improve the usability of the output for land administration [14]. The actual situation may be that the necessary cadastral or field-boundary data are unavailable or inaccurate after the consolidation and terrace abandonment work, etc. Hard parcel aggregation may also occur within-field heterogeneity in the early abandonment stage, when only a portion of a field is overgrown. The previous studies have shown that a good model should be pixel-level flexible and retain parcel-level consistency in its representation, and no external post-processing rules are required.

2.2. Deep Feature Fusion for Agricultural Segmentation

Encoder-decoder networks remain widely used in agricultural segmentation because skip connections retain location detail, but their local continuity does not by itself resolve management ambiguity [15]. Transformer encoders can extend spatial interaction and have improved crop and field extraction, although window partitioning and limited labelled samples may weaken narrow-boundary stability [16]. Feature pyramids address a different requirement by combining high-resolution texture with deep semantic context [17]. The unresolved issue is therefore not simply whether several scales are available, but whether wide parcel context can be fused without overwhelming shallow boundaries and whether unreliable sensor channels can be downweighted when seasonal observations are incomplete. For abandoned-cropland mapping, this interaction matters because a field-scale management pattern and a one-pixel ridge may be simultaneously diagnostic.

Large-kernel convolutional backbones provide a convolutional route to broad spatial support through efficient factorization or reparameterization [18]. Channel recalibration is complementary because optical indices, radar responses, terrain variables, and parcel context do not remain equally informative across seasons or landscapes. Existing work leaves a task-specific gap at the intersection of wide-receptive-field encoding, sensor-aware channel weighting, and boundary-preserving scale fusion. The present study addresses this gap by testing whether their coordinated use improves mean segmentation accuracy, class-specific overlap, field-edge agreement, regional transfer behavior, and confidence calibration. Attention values are treated as conditional responses of the trained model, not as causal explanations of land-management status.

3. Proposed UniRepLKNet SE Framework

3.1. Multisource Input and Architecture Overview

The model receives a stacked feature cube composed of seasonal Sentinel-2 reflectance and vegetation or moisture indices, Sentinel-1 VV and VH backscatter, slope and elevation, and parcel-context distance maps. Optical variables, radar variables, terrain attributes, and contextual distances are normalized within their own sensor groups using training-subset statistics; validation and test statistics are never used for normalization. This grouping prevents the higher variance of radar backscatter from dominating the initial filters while retaining physically meaningful differences within each source. UniRepLKNet forms the backbone, with large-kernel depthwise convolutions for spatial aggregation and pointwise projections for channel interaction. The shallow stage retains 10 m field margins, whereas deeper stages expand support to whole parcels and neighboring management patterns. Figure 1 connects the multisource cube, the four large-kernel stages, the SE recalibration units, the cross-scale decoder inputs, and the final abandoned-cropland probability map.

The relative strides of the four stages in the encoder for the input grid are 4, 8, 16 and 32. A small effective kernel is used in the first stage to avoid blurring the fine terrace ridges, and in the following stages, after feature map downsampling, the spatial support has been expanded. This schedule is more efficient than using the largest kernel at full resolution, and it gradually accumulates local edge descriptions into parcel-group contexts. The width of the stage should be large enough to display the representations of the two middle rows, and recognizable field shapes and sufficient adjacent land need to be maintained to distinguish organized abandonment from isolated fallow. All of these stages are connected to the decoder, and otherwise, the necessary missing evidence at various resolutions for different field sizes would be lost.

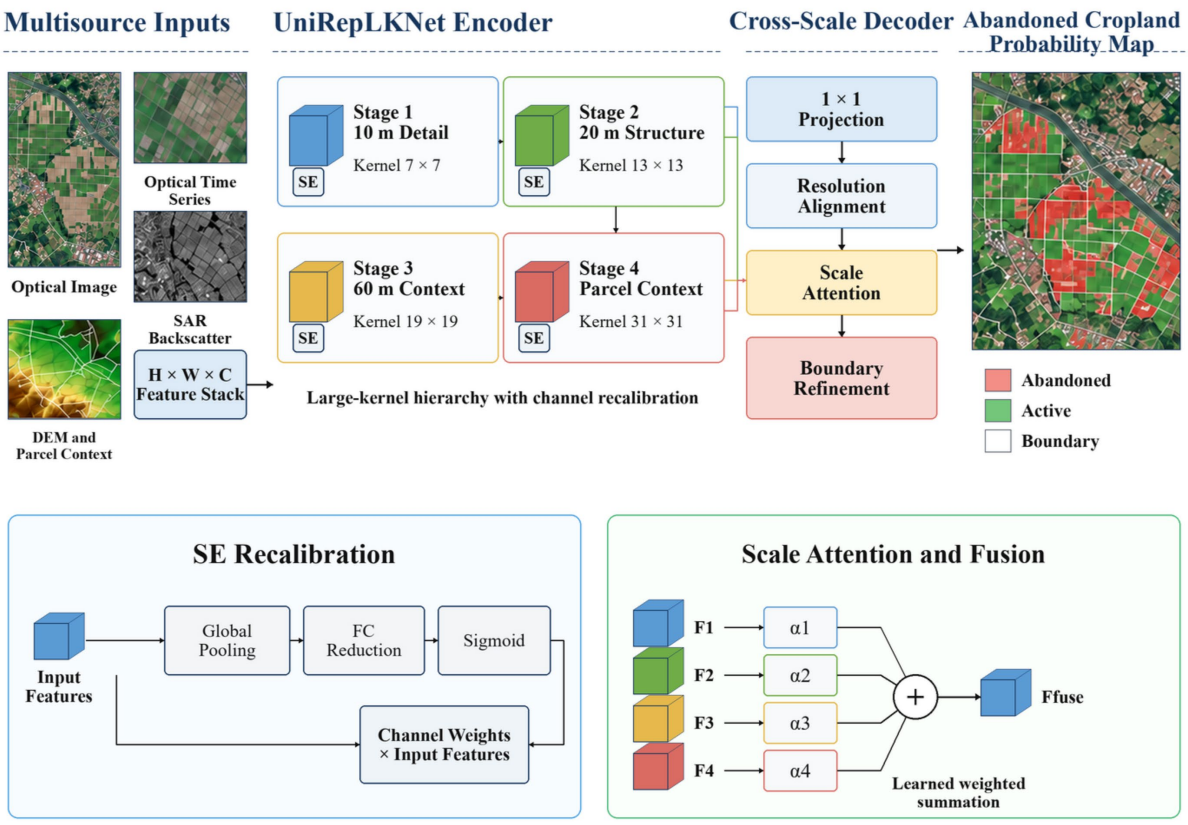


Figure 1. UniRepLKNet SE architecture with multisource inputs, large kernel blocks, SE recalibration, decoder fusion, and abandoned cropland output

3.2. Large Kernel Representation

For an input tensor $X \in \mathbb{R}^{H\times W\times C}$, the stem passes its output through a sequence of large-kernel residual blocks. Wide kernels are suited to agricultural parcels because abandonment is often expressed as a field-level pattern rather than as a single local texture. The stage- s representation is denoted by F_s, and its spatial size decreases as semantic abstraction increases. The following equation formalizes the residual transformation by combining depthwise spatial aggregation with pointwise channel projection.

$$F_s = X_s + PW_2\left(GELU\left(DW_k\left(PW_1(X_s)\right)\right)\right) \tag{1}$$

In this equation, DW_k is the depthwise operator with kernel size k, whereas PW_1 and PW_2 are pointwise projections that control channel dimensionality. The residual path protects local edge evidence, and the wide spatial operator captures row patterns, uncultivated-field continuity, and diffuse vegetation structure. Kernel size increases from shallow to deep stages so that early features retain boundaries and later features integrate evidence from entire parcel groups. Residual normalization is placed before the depthwise operator to stabilize optimization when radar and optical channels have different response distributions. Depthwise aggregation keeps the parameter increase moderate because every channel is filtered independently before pointwise mixing. The block can therefore enlarge spatial support without imposing the quadratic interaction cost associated with dense global attention. During deployment, its regular convolutional structure also supports predictable tiled inference and avoids window-partition artifacts near parcel edges.

3.3. SE Channel Recalibration

The SE module is inserted after each large-kernel transformation to recalibrate channels from a global descriptor. Scale-specific placement is important because the most informative sensor group changes with season and landscape: radar may dominate under cloud contamination, red-edge indices may be decisive during regrowth, and terrain variables may become informative in marginal uplands. Global context is condensed through the pooling equation below, in which z_c summarizes channel c over the complete $H \times W$ feature plane.

$$z_c = \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W} F_c(i,j) \tag{2}$$

The descriptor is deliberately compact, ensuring that recalibration reflects broad scene evidence rather than introducing another spatial classifier. The resulting descriptor enters the channel-recalibration equation, where two lightweight projections and a nonlinear activation produce bounded channel weights.

$$F_c' = F_c \odot \sigma\big(W_2\delta(W_1 z_c)\big) \tag{3}$$

As expressed in the equation, the sigmoid function bounds the learned weights and the element-wise product rescales each feature channel without deleting its residual information. A reduction ratio of four is used in the gating multilayer perceptron to keep the additional parameter cost small. In abandoned-cropland mapping, the operation suppresses bright soil or temporary low-greenness responses when other channels indicate active management, while strengthening persistent vegetation disorder and radar roughness changes that might otherwise be averaged out. Large kernels therefore determine where context is collected, whereas the SE units decide which channel responses should receive emphasis within that context.

4. Cross Scale Fusion and Training Protocol

4.1. Scale Attentive Feature Fusion

The decoder merges four encoder scales rather than relying on a single skip-connection path. Each stage feature is projected to a common channel width and resized to the target decoder resolution. Scale attention then balances high-resolution boundary evidence against low-resolution semantic context. Figure 2 presents the complete workflow, including multisource preprocessing, scale-specific encoding, attentive fusion, supervised optimization, validation, and regional map production.

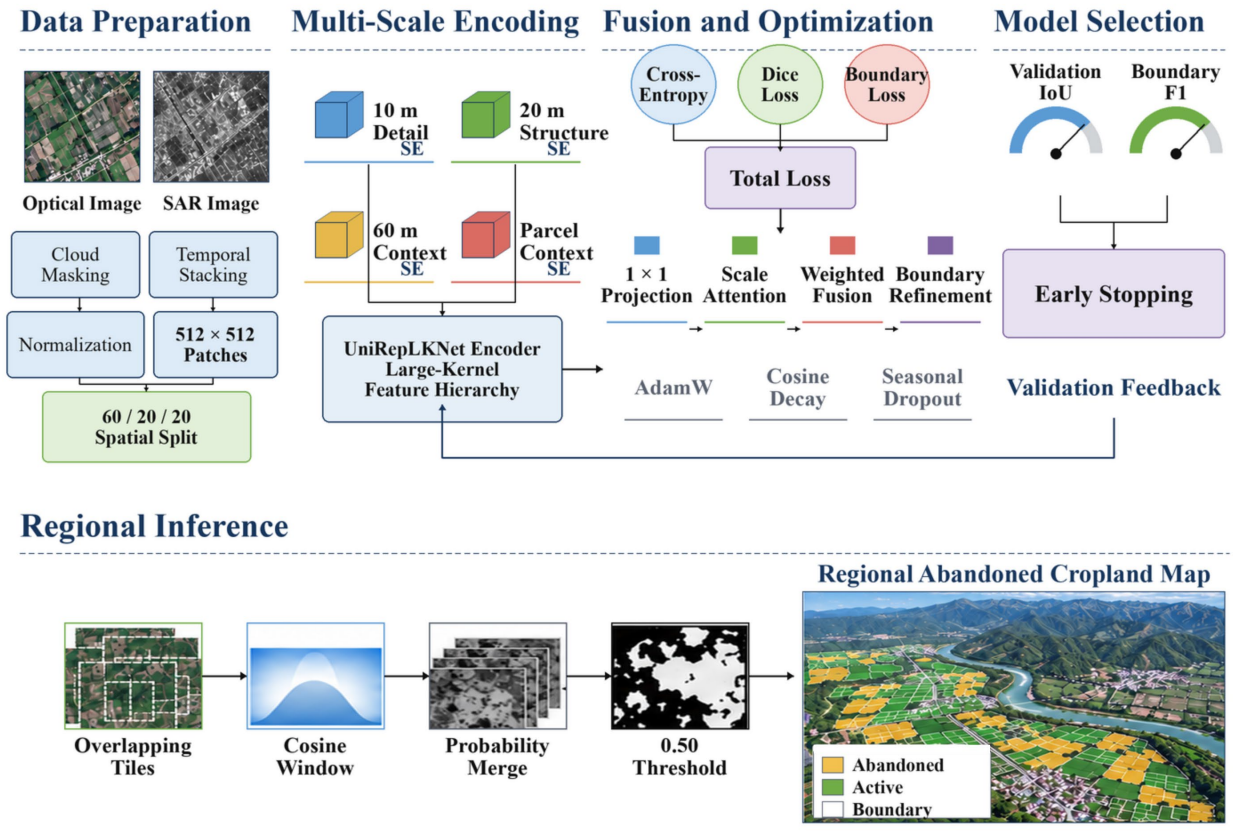


Figure 2. cross scale feature fusion and training workflow from multisource preprocessing to regional abandoned cropland map production

The fused tensor G is constructed from projected features P_s across S scales. A learnable score q_s is calculated for every scale, and softmax normalization determines its relative contribution at the current decoder level. According to the fusion equation, the normalized weights sum to one, preventing any unconstrained branch from increasing the fused magnitude while allowing small parcels and broad abandonment zones to favor different spatial supports.

$$G = \sum_{s=1}^{S} \alpha_s Resize(P_s), \alpha_s = \frac{exp(q_s)}{\sum_{r=1}^{S} exp(q_r)} \tag{4}$$

After weighted summation, a 3×3 refinement convolution restores local consistency and a residual connection carries the highest-resolution projection forward. Fusion is repeated during progressive upsampling rather than being delayed until the final grid, allowing coarse context to guide reconstruction before the output mask is produced. The retained residual protects narrow boundaries even when a coarse branch receives high semantic weight. Attention weights are stored during validation and later related to parcel size and boundary complexity; they are interpreted as conditional model responses rather than as causal importance.

4.2. Composite Training Objective

The training objective combines cross-entropy, Dice loss, and a boundary term. Cross-entropy supports pixel-wise class separation, Dice loss limits the effect of target-class imbalance, and boundary loss penalizes disagreement close to field edges. The composite objective is written in the following equation, where λ_d and λ_b regulate the contributions of overlap and boundary supervision.

$$\mathcal{L} = \mathcal{L}_{ce} + \lambda_d \mathcal{L}_{dice} + \lambda_b \mathcal{L}_{boundary} \tag{5}$$

The boundary term is evaluated in a three-pixel neighborhood around the morphological gradient of the reference mask. The fixed objective uses unit weights for cross-entropy and Dice loss and a boundary coefficient of 0.4; these values are selected on the validation subset and then held constant. Cross-entropy supports pixel separation, Dice loss counters target-class imbalance, and the boundary term discourages excessive smoothing of terraces, irrigation ditches, and road-adjacent plots. Boundary pixels remain in the area losses, so interior classification and edge placement are optimized jointly rather than sequentially. Comparisons share the same spatial split, input variables, augmentation budget, stopping rule, and output resolution, while architecture-specific prediction heads remain unchanged. A sigmoid classifier produces the probability grid. The only post-processing removes connected regions smaller than four pixels when they have no support from parcel-context features; no contour smoothing or parcel snapping is applied.

4.3. Optimization and Regional Inference

Training uses 256×256 pixel patches with random rotation, spectral jitter, radar-noise injection, and seasonal dropout. Seasonal dropout removes one optical date at random and is interpreted as a robustness perturbation, not as a complete physical model of correlated cloud gaps. Target-aware sampling increases the probability of patches containing early abandonment, difficult active-cropland negatives, or field boundaries without duplicating validation or test objects. AdamW optimization uses a five-epoch warm-up followed by cosine decay, a batch size of eight, mixed precision, and gradient clipping. Training is capped at 100 epochs and stops after 12 epochs without improvement in the joint validation criterion. Model selection combines validation abandoned-cropland IoU with boundary F1 because high area accuracy with unstable field limits is unsuitable for inventory updates.

Implementation choices remain conservative for regional production. Group normalization replaces batch normalization in the deepest stages because the multisource cube limits practical batch size. During inference, neighboring tiles overlap by 128 pixels and are merged through a cosine window whose weights decline toward tile edges. Accumulated probabilities are divided by accumulated weights before thresholding, eliminating tile seams without a separate smoothing model that could alter narrow parcel boundaries. No test-time augmentation is applied in the reported configuration. Validation uses the same overlap-merge procedure as testing, so early stopping reflects the complete mapping pipeline rather than isolated patch predictions.

5. Experimental Data and Analysis

5.1. Study Data and Experimental Configuration

The controlled benchmark contains 18,420 labelled image patches stratified into alluvial plain, hilly terrace, peri-urban fringe, and rainfed plateau conditions. These strata represent contrasting agricultural landscapes within the internally consistent evaluation design and should not be interpreted as an independently released geographic survey. Labels cover active cropland, abandoned cropland, woodland and shrub, built-up land, water, and bare or construction soil. Early abandonment, degraded terraces, and partial volunteer vegetation are included, whereas a single season of low greenness is not sufficient for an abandonment label. The dataset contains 6,210 patches with at least one abandoned-cropland object, 7,460 patches dominated by active cultivation, and 4,750 non-cropland patches. Spatial blocking is used to limit optimistic error from nearby landscape samples [19]. Whole field objects and their immediate contextual neighborhoods are assigned to one block before the 60/20/20 training, validation, and test split, preventing adjacent fragments of the same parcel from entering different subsets.

Table 1 shows the composition of the data and sensor variables in the experiment. The subset of alluvial plains has 4,920 patches and a high concentration of active fields; at the same time, the 4,310-patch hilly terrace subset has an abandoned-pixel ratio of 29.7%. The construction soil and idle plots at the peri-urban fringe are difficult to control, and

the canopy development and irregular seasonal response in the rainfed plateau are sparse. Previous phenology-based studies have shown that drought-induced low vigor can resemble agricultural abandonment in the absence of temporal evidence [20]. Therefore, radar observations, terrain attributes and deliberately selected low-vigor active fields were included in the study to reduce the influence of optical greenness.

Table 1. Dataset composition and feature groups used for abandoned cropland mapping

Zone	Patches	Abandoned Pixels	Dominant Challenge	Feature Groups
Alluvial plain	4,920	18.4%	Short fallow and crop residue	Optical, radar, parcel distance
Hilly terrace	4,310	29.7%	Small fields and steep slopes	Optical, radar, DEM, slope
Peri-urban fringe	4,760	16.9%	Idle construction soil	Optical, radar, road distance
Rainfed plateau	4,430	23.1%	Drought and sparse canopy	Optical, radar, terrain

The class-frequency profile in Figure 3(a) shows that active cropland accounts for 43.6% of labeled pixels, while abandoned cropland contributes 21.8%; the resulting 2.0:1 imbalance is substantial but does not reduce the target class to a rare residual category. Figure 3(b) separates the phenological distributions: the median NDVI amplitude is 0.54 for active cropland and 0.31 for abandoned cropland, yet the highlighted 0.25-0.38 overlap band indicates why a fixed temporal threshold would confuse weak crops, bare soil, and early abandonment. Terrain heterogeneity is visible in Figure 3(c), where the upper quartile of terrace slope reaches 18.7 degrees, more than twice the 7.8-degree value in the peri-urban fringe; this spread test whether the encoder can preserve narrow, topographically distorted parcels. Figure 3(d) shows that valid optical observations fall from 91% in March to 58% in July, whereas radar availability remains between 93% and 96% throughout the year. The 35-percentage-point July gap provides a direct test of whether channel recalibration can shift reliance toward radar when optical coverage deteriorates.

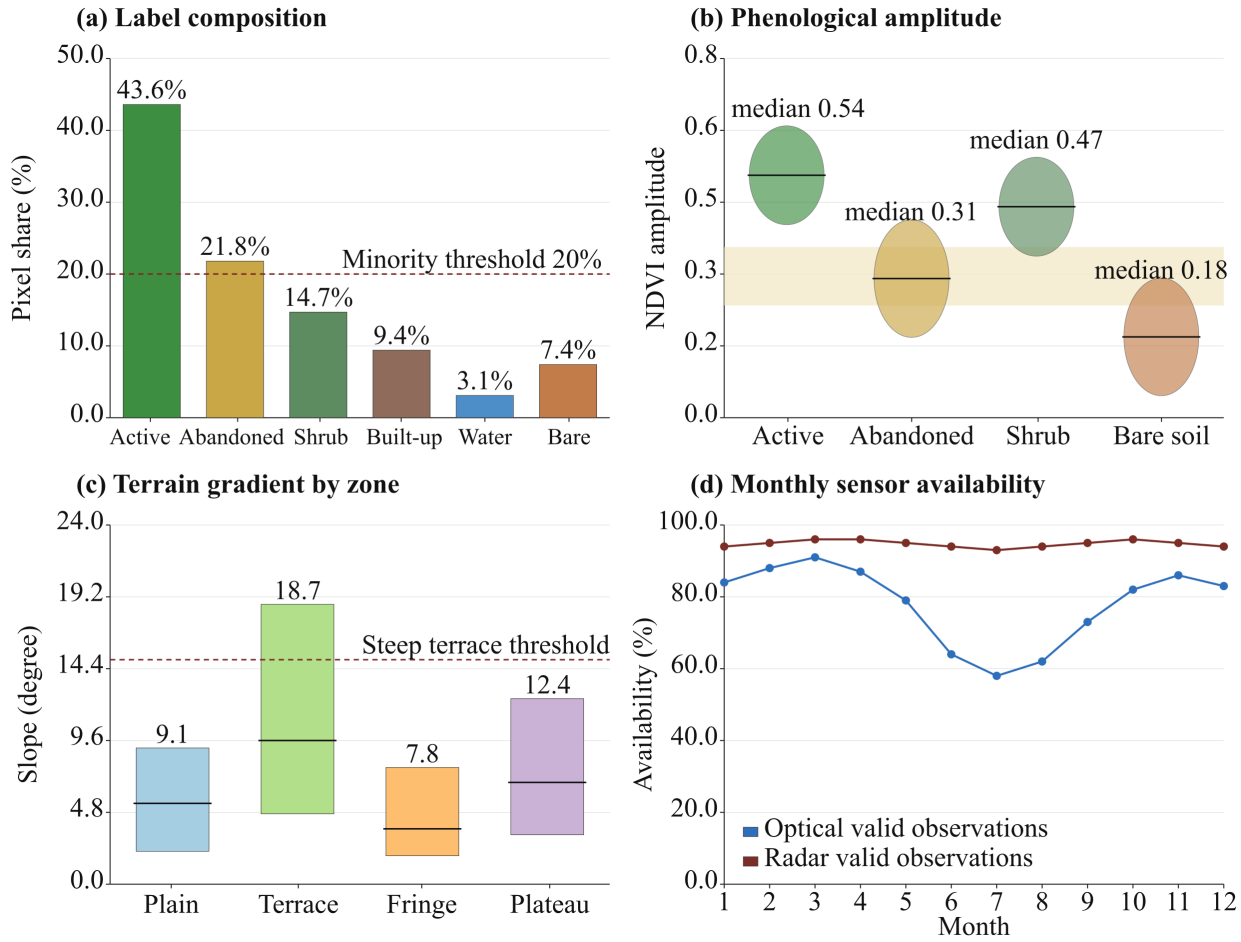


Figure 3. Dataset diagnostics for abandoned cropland mapping: (a) label composition; (b) phenological amplitude; (c) terrain gradient by zone; and (d) monthly sensor availability

Quality control is performed at patch and object levels. Patches with more than 35% cloud cover after masking are excluded, and remaining cloudy pixels are interpolated only when at least two neighboring dates contain valid observations. Abandonment labels require convergent evidence from seasonal composites, radar-backscatter profiles,

parcel geometry, and high-resolution basemap interpretation; isolated low greenness is not accepted because it may indicate fallow, drought stress, harvest, or land-use transition [21]. Ambiguous objects are retained only after consensus interpretation finds no continuous evidence of sowing or harvesting. Direct field verification is not available for every object, which limits the labels to remote-sensing evidence rather than a legal determination of management status. Evaluation therefore reports pixel-level overall accuracy and mean F1, abandoned-cropland IoU, boundary F1, object recovery, false-positive rate, and calibration error so that a high aggregate score cannot conceal edge displacement or landscape-specific failure.

5.2. Comparative Accuracy and Spatial Error Analysis

UniRepLKNet-SE is compared with U-Net, DeepLabV3+, HRNet-OCR, Swin-UNet, and a UniRepLKNet variant without SE recalibration or scale attention. All methods use the same spatial blocks, input variables, augmentation exposure, optimization budget, early-stopping criterion, and validation-selected 0.50 decision threshold; architecture-specific encoders and prediction heads are retained. Such controls are necessary because augmentation and threshold selection can obscure architecture-level differences [22]. Table 2 reports point estimates on the untouched blocked test set: UniRepLKNet-SE reaches 93.8% overall accuracy, 89.6% mean F1, 78.4% abandoned-cropland IoU, and 76.5% boundary F1. Relative to Swin-UNet, the gains are 1.3, 4.8, and 5.3 percentage points, respectively. The false-positive rate is 8.7%, compared with 12.9% for DeepLabV3+, a 32.6% relative reduction. These differences describe this benchmark and are not presented as universal effects across sensors or regions.

Table 2. Comparative segmentation performance on the blocked test set

Method	OA (%)	mF1 (%)	Abandoned IoU (%)	Boundary F1 (%)	False Positive (%)
U-Net	89.4	83.1	68.9	64.7	15.6
DeepLabV3+	90.7	85.0	71.1	66.8	12.9
HRNet-OCR	91.8	86.7	73.4	69.5	11.7
Swin-UNet	92.5	87.9	73.6	71.2	10.8
UniRepLKNet	92.9	88.4	75.2	72.9	10.1
UniRepLKNet-SE	93.8	89.6	78.4	76.5	8.7

The spatial distribution of residual mistakes is shown in Figure 4(a), where the normalized error-density surface peaks at 0.42 in the peri-urban cluster and a weaker secondary cluster appears in terrace margins. This pattern indicates that errors concentrate around ambiguous land-use transitions rather than spreading uniformly across the mapped region. Figure 4(b) reports abandoned-cropland IoU values of 81.6% for hilly terrace, 79.1% for rainfed plateau, 77.3% for alluvial plain, and 72.8% for peri-urban fringe. The 8.8-point terrace-to-fringe range is consistent with confusion between rectangular idle construction soil and recently uncultivated fields, a known difficulty when geometric and spectral cues point to different land-use histories [23]. Figure 4(c) shows that the proposed precision-recall curve remains above 0.85 precision until recall reaches approximately 0.82, whereas the comparison curves begin to decline earlier. Figure 4(d) places the median boundary offset at 1.7 pixels for UniRepLKNet-SE, compared with 2.4 pixels for Swin-UNet and 2.8 pixels for DeepLabV3+; 1.7 pixels corresponds to about 17 m at the working grid and falls below the two-pixel operational tolerance.

The error pattern shows the functions of the three architectural components. U-Net generally divides the irregular abandoned plots into small sections because its local filters are very sensitive to changes in weeds and residues in the field. Swin-UNet has improved continuity, but some masks cross the terrace ridge or extend beyond the shrub margin when window-level context dominates a weak boundary. UniRepLKNet-SE has extended parcel continuity with a large-kernel convolution and reduced the impact of optical dates affected by clouds or weak seasonal contrast through channel recalibration. Parcel-aware accuracy studies have shown that, although there are high pixel-level scores, boundary displacement and object fragmentation are still operationally significant [24]. In the current experiment, the gain of the 3.6-point boundary-F1 over the standard UniRepLKNet baseline exceeds the 0.9-point gain in overall accuracy; thus, the added module appears to focus more strongly on improving spatial organization than merely enhancing easy-to-process interior pixels.

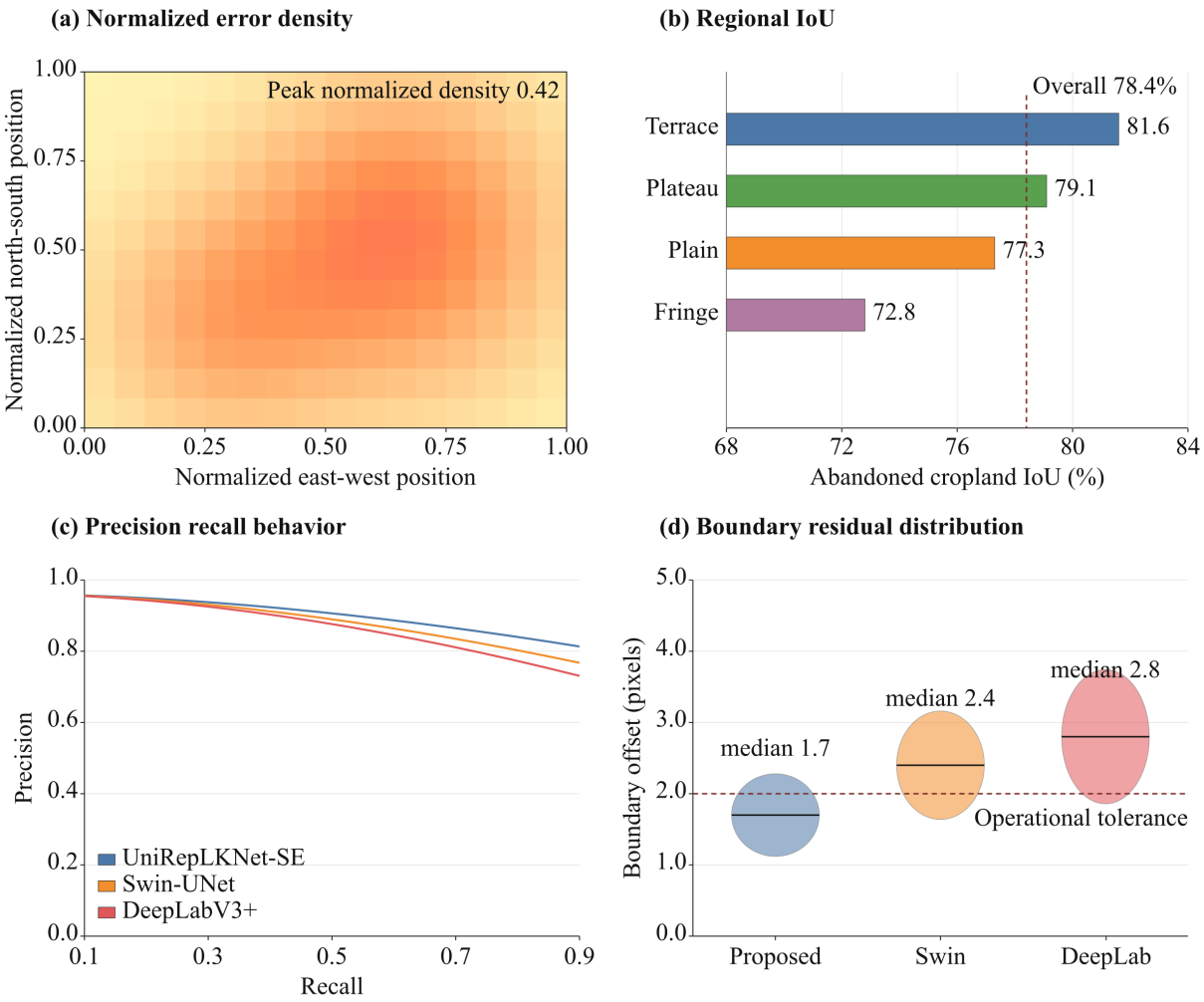


Figure 4. Comparative spatial error analysis: (a) normalized error density; (b) regional abandoned-cropland IoU; (c) precision-recall behavior; and (d) boundary residual distribution

Object-level inspection provides a complementary view. Among abandoned-cropland objects larger than 0.5 ha, UniRepLKNet-SE detects 91.3% with at least 50% mask overlap, compared with 86.9% for Swin-UNet and 84.1% for HRNet-OCR. For objects smaller than 0.2 ha, the detection rate decreases to 72.4%, but it remains 6.2 percentage points above the next strongest baseline. The size-dependent decrease reveals that one or two displaced 10 m pixels can remove a large fraction of a narrow terrace object even when its center is correctly identified. Mean object-area bias is 4.6% for UniRepLKNet-SE, while DeepLabV3+ overestimates area by 11.8% as masks extend into adjacent low-vigor active fields. Across 1,000 object-level bootstrap samples, the proposed model has a 77.1-79.6% IoU interval, compared with 72.0-75.0% for Swin-UNet, and its regional IoU standard deviation is 3.2 points rather than 4.1 points.

5.3. Ablation, Robustness, and Scale Contribution

Controlled removals examine the performance associated with large kernels, SE recalibration, cross-scale attention, and boundary supervision. Channel selection is relevant to multisource remote sensing because clouds, moisture, terrain, and season alter sensor reliability [25]. Replacing large kernels with 3 by 3 blocks reduces abandoned-cropland IoU from 78.4% to 72.6%; removing SE lowers it to 76.1% and increases false positives in sparse active cropland by 13.6%. Simple concatenation without scale attention reaches 76.8% IoU, while removing boundary supervision reduces boundary F1 from 76.5% to 74.1%. All variants use the same optimization and fixed decision threshold. Because the study uses single-component removals rather than a complete factorial design, these differences should be interpreted as conditional module effects within the assembled network, not as independent causal contributions that necessarily add linearly.

Figure 5(a) shows validation IoU rising rapidly during the first 40 epochs, reaching 77.6% near epoch 70, and stabilizing at 78.4% around the epoch-84 stopping point. In Figure 5(b), seasonal-dropout training limits the largest missing-month loss to 2.1 percentage points, whereas the model trained without dropout falls to 73.5% when June is withheld. Random temporal masking can reduce dependence on a single acquisition [26], although this perturbation does not reproduce every correlated cloud-gap pattern. Figure 5(c) reports expected calibration error from 15 equal-

frequency confidence bins: it decreases from 0.091 without SE to 0.044 with SE after model and threshold selection on validation data. Figure 5(d) shows the 10 m attention share changing from 0.44 for small parcels to 0.21 for large parcels, while the 60 m share rises to 0.33. These are conditional responses, not causal importance scores. Figure 5(e) presents controlled IoU differences of 5.8, 2.3, 3.2, and 1.5 points for the large-kernel, SE, fusion, and boundary-supervision removals. Figure 5(f) shows optical importance decreasing from 0.39 to 0.31 under cloudy conditions as radar importance increases from 0.19 to 0.24 and terrain importance from 0.14 to 0.17.

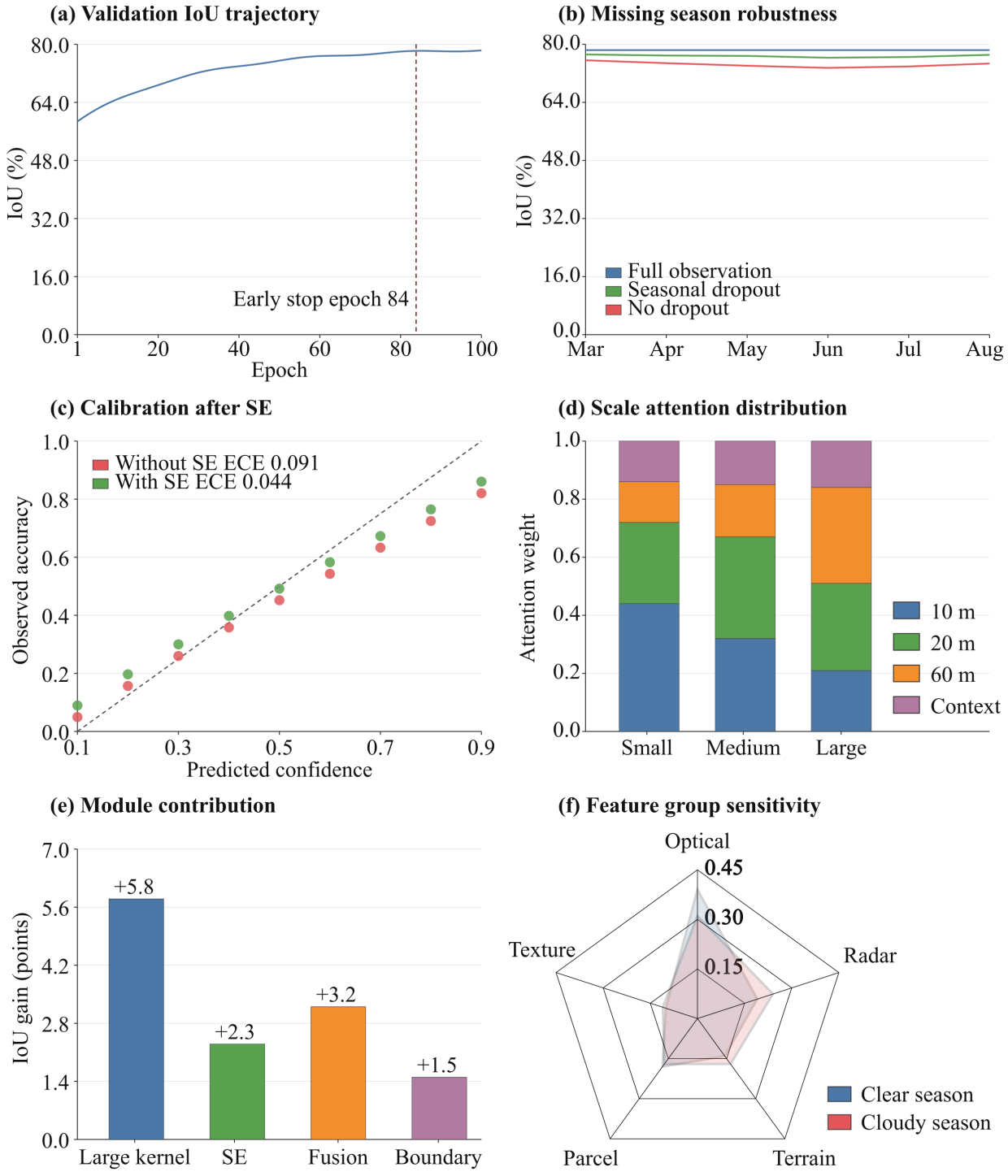


Figure 5. Ablation, robustness, and sensitivity analysis: (a) validation IoU trajectory; (b) missing-season robustness; (c) calibration after SE recalibration; (d) scale-attention distribution; (e) module contribution; and (f) feature-group sensitivity

Scale contribution is examined jointly with parcel geometry in Figure 6. Parcels smaller than 0.35 ha have a mean 10 m attention weight of 0.44 and are concentrated at higher boundary-complexity values, where fine features are needed to retain narrow ridges and tracks. Above 1.20 ha, the combined 20 m and 60 m weight reaches 0.51, and the trend line becomes less sensitive to local boundary oscillation. Bubble size also shows that confidence is lower in highly complex parcels even when fine-scale attention is strong, indicating that the fusion module does not equate spatial detail with

certainty. Cross-region studies commonly report performance loss when field size and terrain distributions change together [27]. In the alluvial-to-terrace transfer test, decoder-only fine-tuning for ten epochs recovers 3.7 IoU points, while full retraining adds only another 0.8 point, suggesting that most transferable information remains in the large-kernel encoder and that scale weighting requires the larger local adjustment.

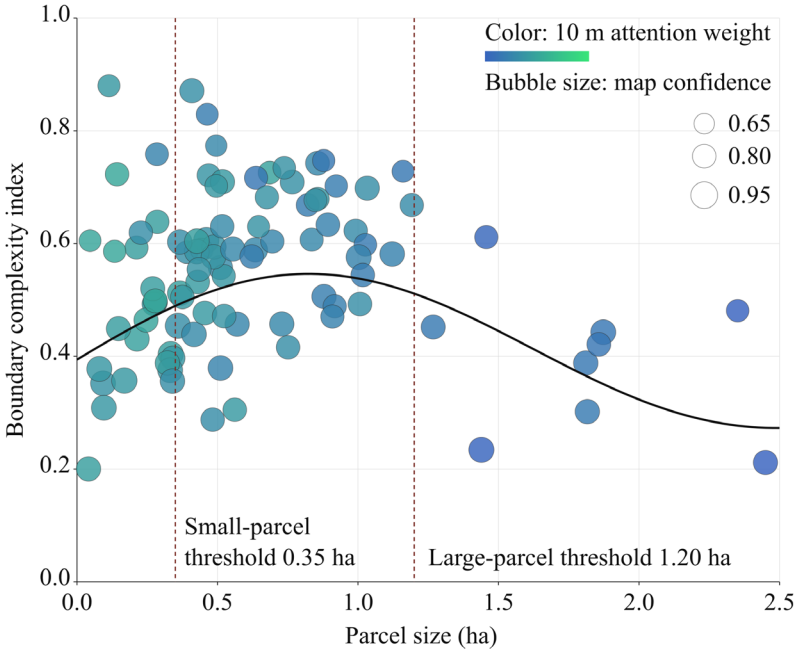


Figure 6. Cross-scale contribution under parcel size and boundary complexity, with color encoding 10 m attention weight and bubble size representing map confidence

Residual errors concentrate in land-use states that provide contradictory cues. Peri-urban idle land is a documented source of confusion because rectangular geometry may indicate former cultivation while exposed soil reflects construction activity [28]. Among 240 manually inspected false-positive objects, 38% are idle construction soil, 27% drought-stressed active fields, 21% shrubland margins, and 14% mixed bare soil and residue. False negatives occur mainly in small terrace parcels under shrub shadow (41%), low-contrast rainfed plots (33%), road-adjacent fields (18%), and narrow strips surrounded by active cultivation (8%). A threshold of 0.50 is selected on validation data and fixed before test evaluation. The test sensitivity check from 0.35 to 0.65 shows less than 1.1-point IoU variation between 0.46 and 0.56; it is diagnostic rather than a second selection step. Under the reported overlapping-tile and mixed-precision configuration, inference reaches 18.6 square kilometers per minute and uses 7.4 GB for eight tiles, compared with 15.9 for Swin-UNet and 20.3 for the unmodified backbone. These throughput values are configuration-specific because complete hardware and software metadata are unavailable.

6. Conclusion

UniRepLKNet-SE was developed and evaluated in this study to introduce cross-scale feature fusion; coordinated large-kernel context modelling, channel recalibration, and multiresolution decoding improved both the thematic accuracy and parcel-boundary quality of abandoned cropland maps.

1. UniRepLKNet-SE achieved 93.8% overall accuracy, 89.6% mean F1, 78.4% abandoned-cropland IoU, and 76.5% boundary F1 on the blocked test set, and limited the false-positive rate to 8.7%. The model found 91.3% of the abandoned-cropland objects larger than 0.5 ha, had a minimum mask overlap of 50%, and its mean object-area bias was 4.6%. As shown in the above results, the model has improved both the accuracy of pixel classification and the spatial completeness and area accuracy required for parcel-level inventories.

2. The performance differences were most pronounced in boundary recovery, difficult negative samples, and incomplete seasonal observations. Relative to Swin-UNet, the proposed model increased abandoned-cropland IoU by 4.8 percentage points and boundary F1 by 5.3 points. Compared with DeepLabV3+, it reduced the false-positive rate from 12.9% to 8.7%. Ablation results attributed 5.8 IoU points to large-kernel representation, 2.3 points to SE recalibration, 3.2 points to cross-scale fusion, and 1.5 points to boundary supervision. Seasonal-dropout training restricted the maximum missing-month IoU loss to 2.1 points, whereas the model trained without dropout decreased to 73.5% IoU when the June observation was removed.

3. The principal innovation is the task-specific coordination of wide spatial context, sensor-channel recalibration, and scale-dependent boundary reconstruction within one convolutional network. Its benefit is clearest when multiseason

multisource observations and reliable parcel-context variables are available. The present evidence remains limited to controlled labelled patches and does not establish multi-year management persistence, legal distinctions between temporary fallow and abandonment, or unrestricted geographic transfer. Errors remain concentrated in peri-urban construction soil, drought-stressed active fields, and small terrace parcels under shrub shadow. Consequently, image evidence should support screening and verification prioritization rather than serve as the sole basis for a legal land-status decision.

The framework provides an engineering basis for regional abandoned-cropland screening; future work should add multi-year temporal reasoning, field and administrative records, externally held regions, uncertainty-guided surveys, and complete hardware and software reporting before policy-scale deployment.

7. Declarations

7.1 Author Contributions

Ioachim Iancu contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision, project administration, and funding acquisition. Laurentiu Jecu contributes to validation, analysis, investigation, data collection, draft preparation. 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 Data 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 ChatGPT-5.5 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] Li, G., Jiang, C., Gao, Y., & Du, J. (2025). Spatial heterogeneity characteristics and driving mechanisms of abandoned farmland in different scales and regions in China. Land Degradation & Development, 36(5), 1452–1466. https://doi.org/10.1002/ldr.5394

[2] Liu, T., Yu, L., Liu, X., Peng, D., Chen, X., Du, Z., Tu, Y., Wu, H., & Zhao, Q. (2025). A global review of monitoring cropland abandonment using remote sensing: Temporal–spatial patterns, causes, ecological effects, and future prospects. Journal of Remote Sensing, 5, Article 0584. https://doi.org/10.34133/remotesensing.0584

[3] Wuyun, D., Sun, L., Chen, Z., Crusiol, L. G. T., Dong, J., Wu, N., Bao, J., Chen, R., Sun, Z., Hasituya, & Zhao, H. (2025). Temporal segmentation method for 30-meter long-term mapping of abandoned and reclaimed croplands in Inner Mongolia, China. International Journal of Applied Earth Observation and Geoinformation, 136, 104399. https://doi.org/10.1016/j.jag.2025.104399

[4] Prishchepov, A. V., Anders, K., Feranec, J., Goga, T., Gradinaru, S. R., Kolář, J., Pazur, R., Potůčková, M., Zagajewski, B., & Kupková, L. (2025). The progress and potential directions in the remote sensing of farmland abandonment. Remote Sensing of Environment, 331, 115019. https://doi.org/10.1016/j.rse.2025.115019

[5] Liu, X., Zhang, J., Duan, Y., & Zhou, J. (2025). CPVF: Vectorization of agricultural cultivation field parcels via a boundary–parcel multi-task learning network in ultra-high-resolution remote sensing images. ISPRS Journal of Photogrammetry and Remote

Sensing, 226, 267–299. https://doi.org/10.1016/j.isprsjprs.2025.05.019

[6] Xu, F., Yao, X., Zhang, K., Yang, H., Feng, Q., Li, Y., Yan, S., Gao, B., Li, S., Yang, J., Zhang, C., Lv, Y., Zhu, D., & Ye, S. (2024). Deep learning in cropland field identification: A review. Computers and Electronics in Agriculture, 222, 109042. https://doi.org/10.1016/j.compag.2024.109042

[7] Gao, Y., Luo, X., Gao, X., Yan, W., Pan, X., & Fu, X. (2024). Semantic segmentation of remote sensing images based on multiscale features and global information modeling. Expert Systems with Applications, 249, 123616. https://doi.org/10.1016/j.eswa.2024.123616

[8] Wang, Y., Feng, L., Sun, W., Wang, L., Yang, G., & Chen, B. (2024). A lightweight CNN-Transformer network for pixel-based crop mapping using time-series Sentinel-2 imagery. Computers and Electronics in Agriculture, 226, 109370. https://doi.org/10.1016/j.compag.2024.109370

[9] Hamidi, M., Homayouni, S., Safari, A., & Hasani, H. (2024). Deep learning based crop-type mapping using SAR and optical data fusion. International Journal of Applied Earth Observation and Geoinformation, 129, 103860. https://doi.org/10.1016/j.jag.2024.103860

[10] Huang, J., Yuan, X., Lam, C.-T., Ke, W., & Huang, G. (2024). Large kernel convolution application for land cover change detection of remote sensing images. International Journal of Applied Earth Observation and Geoinformation, 132, 104077. https://doi.org/10.1016/j.jag.2024.104077

[11] Long, Y., Sun, J., Wellens, J., Colinet, G., Wu, W., & Meersmans, J. (2024). Mapping the spatiotemporal dynamics of cropland abandonment and recultivation across the Yangtze River Basin. Remote Sensing, 16(6), 1052. https://doi.org/10.3390/rs16061052

[12] Liu, X., Wang, S., Zhang, X., Zhen, L., Ma, C., Naing, S. Y., Liu, K., & Li, H. (2025). Monitoring spatiotemporal dynamics of farmland abandonment and recultivation using phenological metrics. Land, 14(9), 1745. https://doi.org/10.3390/land14091745

[13] Brickner, N. Z., Fine, L., Rozenstein, O., & Paz-Kagan, T. (2025). Field crop mapping using machine learning and multi-sensor satellite fusion: Toward dynamic agricultural monitoring. Smart Agricultural Technology, 12, 101650. https://doi.org/10.1016/j.atech.2025.101650

[14] Yan, S., Yao, X., Sun, J., Huang, W., Yang, L., Zhang, C., Gao, B., Yang, J., Yun, W., & Zhu, D. (2024). TSANet: A deep learning framework for the delineation of agricultural fields utilizing satellite image time series. Computers and Electronics in Agriculture, 220, 108902. https://doi.org/10.1016/j.compag.2024.108902

[15] KC, S., Aryal, J., & Ryu, D. (2025). A novel architecture for automated delineation of the agricultural fields using partial training data in remote sensing images. Computers and Electronics in Agriculture, 234, 110265. https://doi.org/10.1016/j.compag.2025.110265

[16] Li, Z., Wang, Y., Tian, F., Zhang, J., Chen, Y., & Li, K. (2024). BAFormer: A novel boundary-aware compensation UNet-like Transformer for high-resolution cropland extraction. Remote Sensing, 16(14), 2526. https://doi.org/10.3390/rs16142526

[17] Huang, X., Wang, H., & Li, X. (2024). A multi-scale semantic feature fusion method for remote sensing crop classification. Computers and Electronics in Agriculture, 224, 109185. https://doi.org/10.1016/j.compag.2024.109185

[18] Liu, B., Li, B., & Li, S. (2025). LK-UNet: Large kernel convolution-driven U-shaped network for semantic segmentation of high-resolution Earth surface images. Advances in Space Research, 75(10), 7020–7034. https://doi.org/10.1016/j.asr.2025.02.033

[19] Stock, A. (2025). Choosing blocks for spatial cross-validation: Lessons from a marine remote sensing case study. Frontiers in Remote Sensing, 6, 1531097. https://doi.org/10.3389/frsen.2025.1531097

[20] Long, Y., Sun, J., Wen, Y., Wang, C., Dong, X., Huang, P., Wu, W., Chen, J., & Ding, M. (2025). Remote sensing approaches for cropland abandonment perception in southern hilly and mountainous areas of China: A review. Smart Agriculture, 7(6), 58–74. https://doi.org/10.12133/j.smartag.SA202505022

[21] Bhatti, M. T., Gilani, H., Ashraf, M., Iqbal, M. S., & Munir, S. (2024). Field validation of NDVI to identify crop phenological signatures. Precision Agriculture, 25, 2245–2270. https://doi.org/10.1007/s11119-024-10165-6

[22] Hadir, A., Adjou, M., Assainova, O., Palka, G., & Elbouz, M. (2025). Comparative study of agricultural parcel delineation deep learning methods using satellite images: Validation through parcels complexity. Smart Agricultural Technology, 10, 100833. https://doi.org/10.1016/j.atech.2025.100833

[23] Karim, M., Guan, H., Zhang, J., & Ayoub, M. (2025). ViT-ChangeFormer: A deep learning approach for cropland abandonment detection in Lahore, Pakistan using Landsat-8 and Sentinel-2 data. Remote Sensing Applications: Society and Environment, 37, 101468. https://doi.org/10.1016/j.rsase.2025.101468

[24] Ma, H., Shen, X., Luo, Z., Chen, P., Guan, B., Zheng, M., & Yu, W.-S. (2024). Extracting agricultural parcel boundaries from high spatial resolution remote sensing images based on a multi-task deep learning network with boundary enhancement mechanism. IEEE Access, 12, 112038–112052. https://doi.org/10.1109/ACCESS.2024.3415169

[25] Shuai, S., Zhang, Z., Zhang, T., Luo, W., Tan, L., Duan, X., & Wu, J. (2024). Innovative decision fusion for accurate crop/vegetation classification with multiple classifiers and multisource remote sensing data. Remote Sensing, 16(9), 1579. https://doi.org/10.3390/rs16091579

[26] Zhang, X., Cai, Z., Hu, Q., Yang, J., Wei, H., You, L., & Xu, B. (2024). Improving crop type mapping by integrating LSTM with temporal random masking and pixel-set spatial information. ISPRS Journal of Photogrammetry and Remote Sensing, 218, 87–101. https://doi.org/10.1016/j.isprsjprs.2024.10.013

[27] Wijesingha, J., Dzene, I., & Wachendorf, M. (2024). Evaluating the spatial–temporal transferability of models for agricultural land cover mapping using Landsat archive. ISPRS Journal of Photogrammetry and Remote Sensing, 213, 72–86. https://doi.org/10.1016/j.isprsjprs.2024.05.020

[28] Feng, S., Jiang, S., Liu, X., Zhang, L., Gan, Y., Xia, N., Wu, W., & Zhou, C. (2025). Extraction of abandoned cropland using multisource remote sensing images in suburban regions: A case study of Zengcheng, Guangdong Province. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 18, 18055–18067. https://doi.org/10.1109/JSTARS.2024.3504518