agriculture

MDPI

Article

Early Mapping of Farmland and Crop Planting Structures Using Multi-Temporal UAV Remote Sensing

Lu Wang [1,2], Yuan Qi [1,*], Juan Zhang [1,2], Rui Yang [1], Hongwei Wang [1], Jinlong Zhang [1,2] and Chao Ma [1,2]

[1] State Key Laboratory of Cryospheric Science and Frozen Soil Engineering, Key Laboratory of Remote Sensing of Gansu Province, Heihe Remote Sensing Experimental Research Station, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou 730000, China; wanglu231@mails.ucas.ac.cn (L.W.); zhangjuan232@mails.ucas.ac.cn (J.Z.); yangrui@lzb.ac.cn (R.Y.); wanghw@lzb.ac.cn (H.W.); zjinlong@lzb.ac.cn (J.Z.); machao23@mails.ucas.ac.cn (C.M.)

[2] University of Chinese Academy of Sciences, Beijing 100049, China

* Correspondence: qiyuan@lzb.ac.cn

Abstract

Fine-grained identification of crop planting structures provides key data for precision agriculture, thereby supporting scientific production and evidence-based policy making. This study selected a representative experimental farmland in Qingyang, Gansu Province, and acquired Unmanned Aerial Vehicle (UAV) multi-temporal data (six epochs) from multiple sensors (multispectral [visible–NIR], thermal infrared, and LiDAR). By fusing 59 feature indices, we achieved high-accuracy extraction of cropland and planting structures and identified the key feature combinations that discriminate among crops. The results show that (1) multi-source UAV data from April + June can effectively delineate cropland and enable accurate plot segmentation; (2) July is the optimal time window for fine-scale extraction of all planting-structure types in the area (legumes, millet, maize, buckwheat, wheat, sorghum, maize–legume intercropping, and vegetables), with a cumulative importance of 72.26% for the top ten features, while the April + June combination retains most of the separability (67.36%), enabling earlier but slightly less precise mapping; and (3) under July imagery, the SAM (Segment Anything Model) segmentation + RF (Random Forest) classification approach—using the RF-selected top 10 of the 59 features—achieved an overall accuracy of 92.66% with a Kappa of 0.9163, representing a 7.57% improvement over the contemporaneous SAM + CNN (Convolutional Neural Network) method. This work establishes a basis for UAV-based recognition of typical crops in the Qingyang sector of the Loess Plateau and, by deriving optimal recognition timelines and feature combinations from multi-epoch data, offers useful guidance for satellite-based mapping of planting structures across the Loess Plateau following multi-scale data fusion.

Keywords: unmanned aerial vehicle (UAV); early identification of planting structures; optimal feature selection; multi-source data; multi-temporal remote sensing data

check for updates

Academic Editor: Juan Arribas

Received: 11 September 2025
Revised: 16 October 2025
Accepted: 20 October 2025
Published: 22 October 2025

Citation: Wang, L.; Qi, Y.; Zhang, J.; Yang, R.; Wang, H.; Zhang, J.; Ma, C. Early Mapping of Farmland and Crop Planting Structures Using Multi-Temporal UAV Remote Sensing. *Agriculture* **2025**, *15*, 2186. https://doi.org/10.3390/agriculture15212186

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Satellite and UAV remote-sensing technologies are instrumental in advancing precision agriculture; however, existing studies provide insufficient support for efficient, high-accuracy UAV-based extraction of crop planting structures and lack a workflow transferable to satellite imagery [1].

Satellite remote sensing technology has been widely applied in agriculture [2]. For example, classification accuracy of early-stage crops can be greatly improved by combining

spatiotemporal features with multi-temporal satellite image data (e.g., Sentinel-2) [3]. Similarly, selecting optimal classification features from multi-temporal remote sensing data can significantly enhance early planting structure extraction [4]. In general, early-stage crop identification accuracy can be improved considerably through optimal feature selection based on multi-temporal remote sensing data. While satellite remote sensing offers wide coverage and strong timeliness, satellites also have limitations, including insufficient spatial resolution, long revisit cycles, limited agility for dynamic retasking, and reduced applicability in complex terrain [2,5].

The UAV Remote Sensing System (UAVRSS) plays an important role in fine-scale identification of crop planting structures and field parcels, monitoring of plant diseases and pests, soil moisture assessment, and fertilization decision-making [6]. For fine-scale extraction of planting structures, UAV multi-temporal remote sensing data have been used to classify early-stage vegetables [7] through Recurrent Convolutional Neural Networks (RCNNs) with attention mechanisms, where temporal data are processed using dynamic time warping. Deep semantic segmentation of UAV multispectral data combined with multi-scale features significantly improves classification accuracy in complex farmland scenarios [8]. In addition, rice monitoring models that fuse texture features with vegetation indices from UAV hyperspectral images show high stability across multi-temporal datasets [9]. The combination of a random forest classifier, texture metrics, and spectral features of segmented objects markedly enhances vegetation classification accuracy [10]. Overall, UAV remote sensing—by virtue of its ultra-fine detail capture, outstanding flexibility and timeliness, and markedly lower cost than satellite data—demonstrates clear advantages for improving agricultural efficiency [2]. These include stronger resilience to cloud interference, substantially higher accuracy in crop planting-structure identification enabled by centimeter-level resolution, and faster data responsiveness, among others [11,12]. However, UAV technology has several limitations: constraints on flight altitude and battery endurance in complex terrain [13], markedly reduced coverage at low altitudes [12], extensive visual line-of-sight (VLOS) regulatory requirements [14], and a lack of standardized sensor calibration procedures [13]. To both overcome the limitations of UAV technology and leverage its high maneuverability in extracting crop planting structures, a method for defining optimal temporal windows and feature combinations is needed, along with an investigation of its reuse/transferability potential. Against this backdrop, recent advances in deep learning provide a complementary pathway to operationalize these goals, linking UAV-derived fine detail with temporally explicit modeling across sites.

Over the past decade, deep learning for agricultural remote sensing has evolved from Convolutional Neural Network (CNN)-dominated static representations to dynamic, temporally and attention-aware representations [15–17]. On multitemporal satellite/UAV data, Long Short-Term Memory (LSTM)/Gated Recurrent Unit (GRU) models directly learn crop phenology trajectories [18], approaches that target the challenges of small-object detection (e.g., crop disease spots in farmland) encompass improvements to attention mechanisms and Feature Pyramid Networks (FPNs) [19], and pixel-set encoders with temporal self-attention further improve field-parcel-scale classification robustness [20]. For semantic segmentation, U-Net and DeepLabv3+ remain strong baselines for high-resolution remote sensing [21,22]; combined with spectral indices and Gray-Level Co-occurrence Matrix (GLCM)-based textures, as well as thermal-infrared and Light Detection and Ranging (LiDAR) structural cues, they substantially enhance crop/parcel segmentation under rugged topography and heterogeneous backgrounds [15,17]. In large-area, long time-series mapping of cropping structures, deep models complement machine learning classifiers, enabling cross-regional mapping and early-season interpretation without field-level labels [22,23]. Nevertheless, key gaps remain: a systematic methodology is still lacking

for defining the optimal observation window–discriminative feature set tailored to specific crop lineages in semi-arid, undulating terrains; moreover, cross-scale reuse from UAV to satellite for fine-grained agricultural remote sensing requires further development. Against this background, in a Loess Plateau test area, we derive reusable optimal windows and key features from multi-temporal UAV data via object-level segmentation and feature selection, and quantitatively compare them with RF/SVM/CNN, thereby providing a foundation for rapid transfer and reuse at larger scales.

The Segment Anything Model (SAM), trained on 11 million images and 1.1 billion masks, enables rapid and precise segmentation of images through zero-shot learning [24]. In contrast, the multi-resolution segmentation (MRS) method requires manual setting of scale parameters to adapt to different ground objects. Recent studies indicate that, compared with traditional multi-scale segmentation methods for high-resolution cropland extraction, SAM (used in an unsupervised/zero-prompt manner) achieves an average IoU improvement of 12–28% and increases recall on texture-complex parcels by 18% [25]. Moreover, improved SAM frameworks yield segmentation results that are significantly superior to traditional approaches, with Dice gains of 0.8–6.2% and Jaccard gains of 4.0–7.8%, showing particular advantages for small objects in complex scenes [26]. On multi-scale datasets, SAM attains an average segmentation accuracy of 89.3%, outperforming traditional methods (average 83.5%) [27]. From the standpoint of processing and computational efficiency across different segmentation methods, SAM, under a zero-shot setting, reduces inference time by approximately 30–40% compared with traditional multi-scale segmentation methods (e.g., algorithms based on multi-scale parameter optimization). On high-resolution UAV data, processing time increases with input size [27]. In farmland-parcel extraction, SAM runs end to end at about 0.05 s per image, whereas traditional methods (e.g., multi-scale GMM) take >0.2 s; the efficiency gain comes from its lightweight encoder–decoder design, though memory usage is higher for full-image segmentation [25]. Given its zero-shot capability and alignment with UAV timeliness, this study applies SAM to segment multi-temporal UAV data and compares it with MRS (multi-resolution segmentation) to establish a time-sensitive segmentation scheme for monitoring crops in complex planting areas.

In this study, an experiment on early detection of complex planting structure categories was conducted. Based on the results, a multi-sensor, multi-temporal UAV remote sensing method was developed for the Huachi Agricultural Demonstration Zone in Qingyang on the Loess Plateau. The objective was to address challenges in early identification within farmland characterized by complex planting environments. Six periods of multi-temporal UAV data (multispectral, thermal infrared, and LiDAR) were acquired and combined with ground survey data to perform early farmland identification. SAM was employed for plot segmentation, after which spectral and texture features were fused. Combined with ground point location statistics, the RF algorithm was used to conduct early fine-scale identification of crop types. Time-window sensitivity was analyzed, feature importance was ranked, and crops at different phenological stages were identified using a sliding time window, enabling optimal feature subset selection.

From these results, the optimal timeline for early identification of different crops in the Demonstration Zone and the corresponding key discriminative feature combinations were determined, forming a reusable workflow and parameter set. Consequently, early precise detection of complex planting structures was achieved based on fine delineation of field parcels. Compared with the previously reported multi-scale segmentation plus Random Forest approach using high-resolution data on the Loess Plateau (90% accuracy) [28], our method achieves an absolute accuracy gain of 2.66 percentage points. This outcome not only provides empirical support for early identification of major crops on the Loess Plateau

but also offers a transferable technical pathway for rapid surveys and dynamic updates of large-scale crop planting structures.

2. Data and Research Area

2.1. Overview of the Research Area

The study area is located in the Chenghao Town Agricultural Demonstration Zone, Huachi County, Qingyang City, Gansu Province, China (36.220–36.233° N, 107.9894–108.0244° E), within the hilly–gully loess tableland of the eastern Loess Plateau, at a mean elevation of approximately 1300 m. The agro-environment is characterized by a semi-arid, temperate continental monsoon climate, with precipitation concentrated in June–September; pronounced intra-annual soil-moisture variability and frequent short-duration convective storms elevate erosion risk. Soils are predominantly loessial, silt-dominated (locally termed Huangmian), with high erodibility [29]; the landform is a dissected residual tableland with ridge–gully morphology, producing strong surface heterogeneity and fragmented field boundaries—conditions well suited to evaluating object-based parcel segmentation methods. Cropping is maize-dominated (the combined planting area of maize and wheat exceeds 50% in eastern Gansu [28]), with widespread promotion of soybean–maize strip intercropping to achieve dual grain–legume production (achieves 20–50% savings in water and land use [30]); a wheat–buckwheat "two-year/three-harvest" rotation is also practiced. Additional crops include legumes, millet, sorghum, minor cereals (e.g., foxtail millet/buckwheat), and vegetables. This crop assemblage covers the principal crop types of the Loess Plateau and provides a realistic, challenging testbed for multi-payload UAV observations and high-precision mapping of cropping structures.

2.2. Data Acquisition and Preprocessing

This study utilized data from six UAV-ground simultaneous observation experiments conducted in Qingyang in 2022. The UAV data are summarized in Table 1. The UAV operated in DJI Pilot 2 (SZ DJI Technology Co., Ltd., Shenzhen, China) automatic flight-path planning mode; once the study extent was specified, both sidelap and forward overlap were configured to 70%. At this overlap, data collection for a single payload took ~25–30 min, remaining within a single flight (one battery cycle). Considering the characteristics of multi-source remote sensing data, a standardized preprocessing workflow was applied in modules.

Table 1. UAV and ground-based experimental data.

Parameter	Description
Experiment time	30 April, 17 May, 14 June, 20 July, 15 August, 20 September, 2022
UAV platform	DJI Matrice 300 RTK UAV platform
Sensor configuration	Dual-gimbal system carrying DJI L1 LiDAR camera, DJI P1 survey camera, and Yusense MS600 Pro multispectral sensor
Data acquisition method	Multi-source remote sensing data were obtained through six synchronized UAV and ground experiments.
Flight area	0.6 km^2
Flight altitude	The constant cruising altitude of all sensors was 100 m
Multispectral data bands	B1 (450 nm), B2 (555 nm), B3 (660 nm), B4 (720 nm), B5 (750 nm), and B6 (840 nm)
Ground resolution	7 cm
Frequency of LiDAR echo	3 times
Point cloud density	540 points/m^2
Number of ground sample points	75 points
Ground object types	Maize, wheat, maize-legume striped intercropping, legumes, sorghum, millet, broom corn millet, buckwheat, vegetables, and non-agricultural farmland
Other ground measurement data	Plant height, soil temperature, and plant greenness

For multispectral data, we first ensured clear, cloud-free conditions during data acquisition. Before each multispectral sortie, the Changguang Yuchen MS600Pro sensor was

calibrated using manufacturer-supplied standard reflectance panels, and the incoming irradiance was normalized with the onboard DLS (downwelling light sensor). Pix4Dmapper was used to process the original data, and radiometric calibration was conducted using onboard illuminance sensor data. This ensured a reflectance retrieval error of $\leq$5%. We performed multi-band data fusion in ArcGIS Pro 3.4 and conducted sub-pixel co-registration of multi-temporal imagery in ENVI 5.6. For the fine registration, we used $\geq$50 evenly distributed control points per scene, achieving a geometric accuracy of RMSE $\leq$ 1 pixel.

For 3D point cloud data, the raw data were processed using a point cloud system with noise suppression and outlier removal, followed by initial registration of multi-period datasets. The Iterative Closest Point (ICP) algorithm was applied to achieve millimeter-level fine registration, with spatial registration RMSE $\leq$ 10 mm.

Simultaneously with UAV data acquisition, field surveys of crop growth were conducted in the Demonstration Zone. An RTK positioning device was employed to assist in marking different planting structures on the ground. A total of 75 ground-truth sample points were collected using a stratified scheme that covered all cropping structures and was uniformly distributed across plots (Figure 1). During labeling, we ensured coverage of all single-crop areas within the study region (so that, during visual interpretation, any parcel could be accurately interpreted using ground truth from parcels of the same planting-structure class). A 1 m $\times$ 1 m quadrat was established at the center of each field plot to collect crop location, crop type, plant height, and phenological information. Crop locations were recorded using RTK positioning. Plant height was measured with a ruler on five randomly selected plants and averaged. Canopy greenness was measured with a chlorophyll meter (HM-YD) at the upper and middle positions of three plants; the mean of these readings was taken as the quadrat-level greenness value.

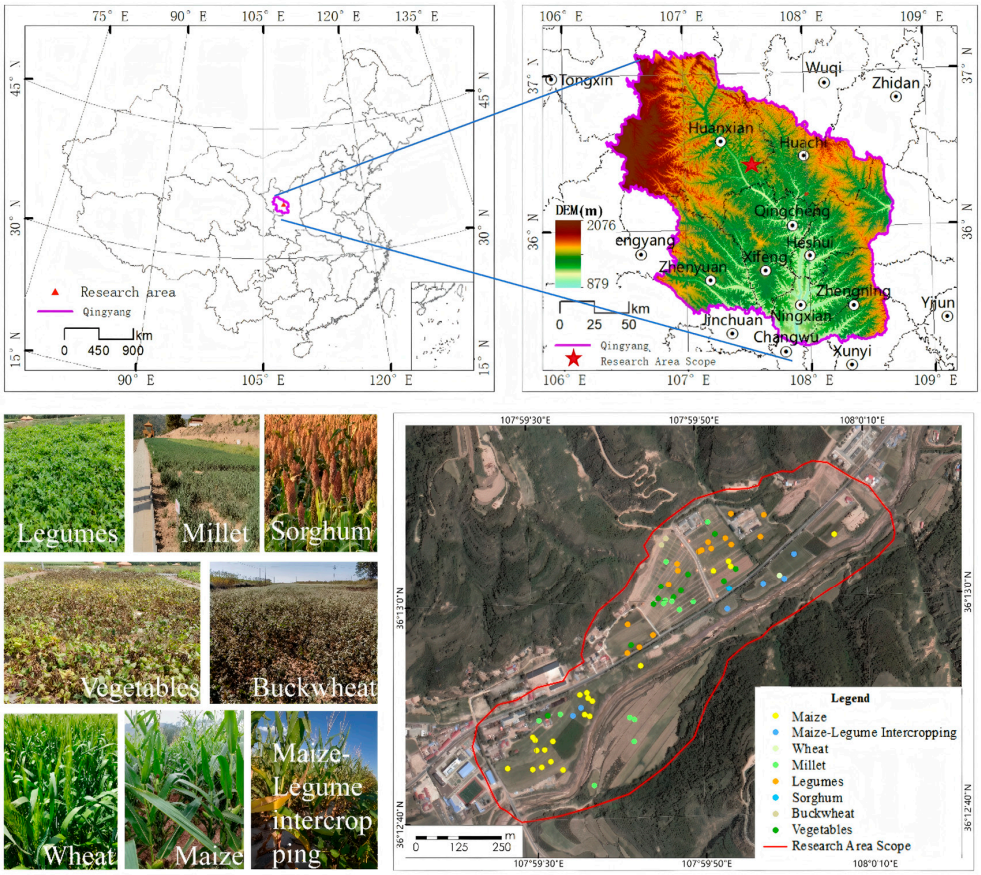


Figure 1. Research area location, sample-point distribution map, and ground photographs of crops.

3. Research Method

3.1. Feature Calculation

The extraction of planting structures using vegetation indices relies on the principle that different planting structures have distinct reflection and absorption characteristics across spectral bands. These differences can be revealed through operations between bands [31]. Meanwhile, there are significant structural differences among the various planting structures in the study area.

In this study, 11 vegetation indices were selected. Given that the UAV multispectral data included six bands, 48 texture features (eight features per band) were generated using the Gray-Level Co-occurrence Matrix (GLCM). The objective was to enhance the distinctiveness of spectral response differences among major crops across growth stages and improve classification accuracy. For texture feature calculation, eight texture values were computed for each band, yielding 48 texture features in total. Texture analysis was conducted using a 3 × 3 window, with direction fixed at 0° and offset values of both X and Y in the spatial correlation matrix set to 1. (The vegetation indices and texture features are displayed in Table 2, where N represents the gray level, P(i,j) represents the co-occurrence probability of gray levels i and j at a specific direction and distance, μx and μy represent the means, and σx and σy represent the standard deviations.)

Table 2. Computation and Detailed Description of Vegetation Indices and Texture Features.

	Feature	Formula	Detailed Description of Features
Vegetation Index	Normalized Difference Vegetation Index (NDVI) [32]	$NDVI = \frac{NIR - Red}{NIR + Red}$	It is sensitive to chlorophyll content and canopy density, enabling early differentiation between germinating crops and bare land.
	Enhanced Vegetation Index (EVI) [33]	$EVI = 2.5 \times \frac{NIR - Red}{NIR + 6Red - 7.5Blue + 1}$	It is sensitive to high Leaf Area Index (LAI) and canopy structure, facilitating early distinction between densely planted and sparsely planted crops.
	Normalized Difference Red Edge Index (NDRE) [34]	$NDRE = \frac{NIR - Red\ Edge}{NIR + Red\ Edge}$	It is sensitive to the nitrogen content and physiological status of early-stage leaves, helping to distinguish legume and millet and crops with significant differences in nitrogen fertilization management.
	Soil-Adjusted Vegetation Index (SAVI) [35]	$SAVI = \frac{NIR - Red}{NIR + Red + L} \times (1 + L)$	It suppresses the interference of early-stage soil background, especially in crop planting areas with a high proportion of bare soil.
	Modified Soil-Adjusted Vegetation Index (MSAVI) [36]	$MSAVI = \frac{2NIR + 1 - \sqrt{(2NIR + 1)^2 - 8(NIR - Red)}}{2}$	It enhances the identification capability in areas with low vegetation coverage.
	Green Normalized Difference Vegetation Index (GNDVI) [37]	$GNDVI = \frac{NIR - Green}{NIR + Green}$	It is sensitive to early chlorophyll content and thus can be used to distinguish early-stage crops with high chlorophyll content.
	Ratio Vegetation Index (RVI) [38]	$RVI = \frac{NIR}{Red}$	It reflects the vegetation amount, enabling early-stage differentiation between bare land and vegetation.
	Simple Ratio Index (SR) [39]	$SR = \frac{NIR}{R}$	It is sensitive to total vegetation amount, assisting in improving accuracy in combination with other indices.

Table 2. *Cont.*

Feature		Formula	Detailed Description of Features
Vegetation Index	Blue Normalized Difference Vegetation Index (BNDVI) [40]	$BNDVI = \frac{NIR - Blue}{NIR + Blue}$	It enhances the early sensitivity to moisture, facilitating early identification of dry land and irrigated crop planting areas.
	Difference Vegetation Index (DVI) [33]	$DVI = NIR - Red$	It reflects the reflectance difference between vegetation and soil and thus is suitable for early identification of sparsely planted crops.
	Modified Non-Linear Vegetation Index (MNLI) [41]	$MNLI = 1.5 \times \frac{NIR^2 - R}{NIR^2 + R + 0.5}$	It is sensitive to spectral changes in medium and high-density vegetation, enabling early identification of densely planted and monoculture crops.
Texture features	Mean (MEAN)	$Mean = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} i \cdot P(i,j)$	It reflects image brightness, facilitating early coarse distinction between densely planted and sparsely planted crop areas.
	Variance (VAR)	$Variance = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (i-\mu)^2 \cdot P(i,j)$	It reflects the discrete degree of spectral values and is sensitive to the spectral inhomogeneity of intercropped vegetation.
	Homogeneity (HOM)	$Homogeneity = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{P(i,j)}{1+\lvert i-j \rvert}$	It reflects the texture uniformity, with high value in densely planted areas and low value in intercropped or vegetable areas.
	Contrast (CON)	$Contrast = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (i-j)^2 P(i,j)$	It reflects the gray-level difference between neighboring pixels, with high values in areas with different canopy height and row spacing.
	Dissimilarity (DIS)	$Dissimilarity = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \lvert i-j \rvert \cdot P(i,j)$	It reflects the average gray-level difference between neighboring pixels, enabling early identification of crops with a high proportion of bare soil between rows.
	Entropy (ENT)	$Entropy = -\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P(i,j) ln P(i,j)$	It reflects the degree of disorder in gray-level distribution, with high entropy values in areas with complex vegetation structures.
	Second Moment (ASM)	$Second\ Moment = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} [P(i,j)]^2$	It reflects the texture regularity, with high values for crops with high uniformity.
	Correlation (CORR)	$Correlation = \frac{\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (i-\mu_x)(j-\mu_y) P(i,j)}{\sigma_x \sigma_y}$	It reflects the degree of correlation between gray-level values in neighboring pixels, with high values for uniformly growing crops and low values for intercropping and vegetables.

Note: Texture features are adapted from [42,43].

3.2. Planting Structure Extraction Based on Optimal Feature Selection

In remote sensing image classification, the object-based method provides strong scale adaptability [44]. Accordingly, the technical route of this study was as follows: farmland was first extracted using fused multispectral, thermal infrared, and LiDAR data. After precise farmland extraction, SAM was employed to segment pixel-based RGB data into object-based monoculture plots. Following farmland segmentation into plots, the RF algorithm was used to select optimal features for both the entire planting structure and monoculture structures based on the features described in Section 3.1. These features

served as the classification basis for early, accurate extraction of planting structures and for defining the timeline for early identification of various planting structures. Finally, the optimally selected features were used for the early, precise extraction of vegetation structures (Figure 2). The objective of this study is to determine, from multi-temporal UAV data, the optimal observation windows and the best discriminative feature combinations for cropland and planting-structure extraction. We hypothesize that object-level segmentation with SAM, coupled with the top-10 features selected by RF, will outperform MRS and CNN/SVM baseline methods; moreover, an April + June observation window can, under time-critical constraints, retain most of the separability achieved in July, enabling earlier identification of cropland and planting structures.

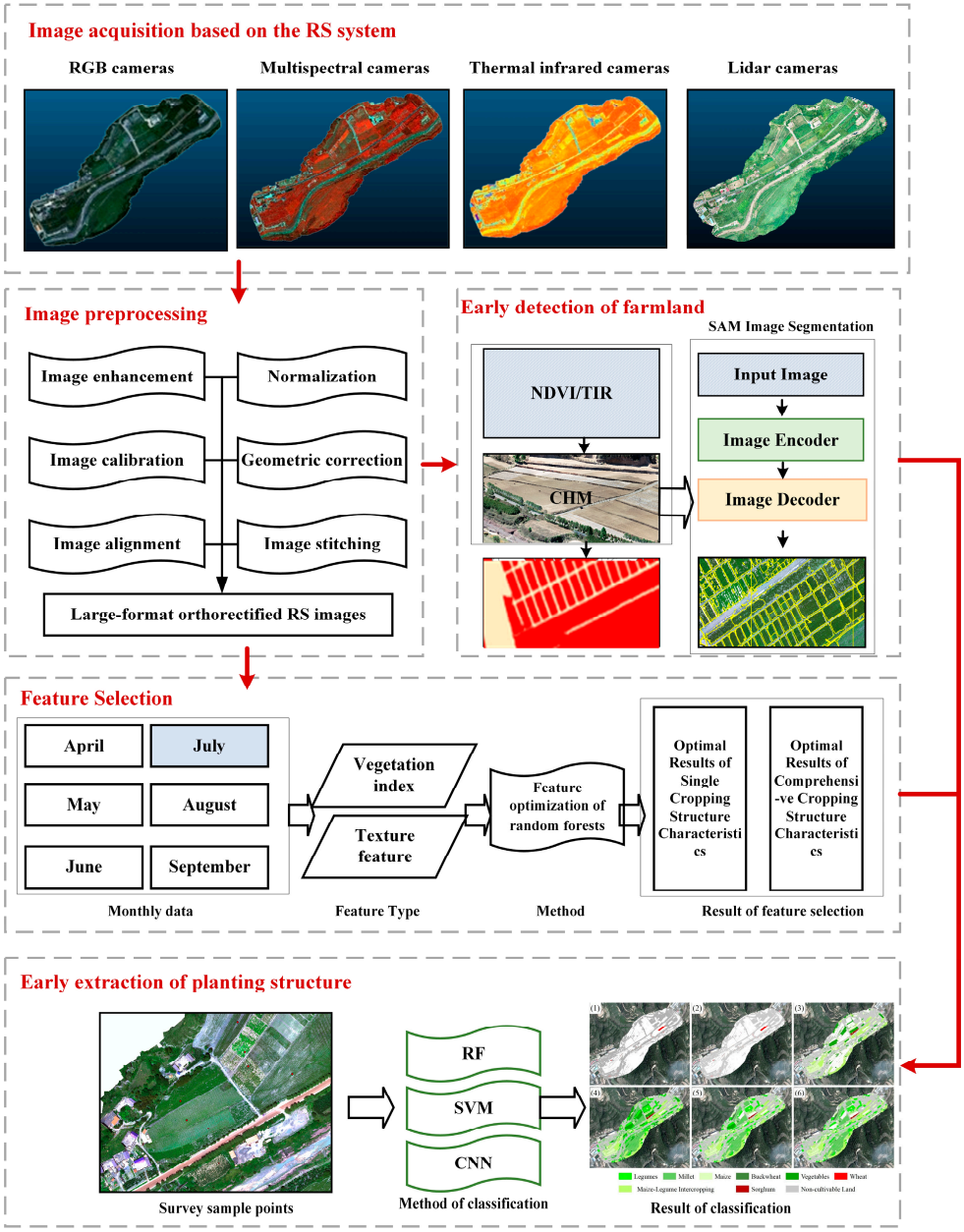


Figure 2. Technical route diagram for fine-scale identification of planting structures and field parcels.

3.2.1. Farmland Extraction

UAV data from three periods (April, May, and June) were used for early farmland identification and precision comparison. Vegetation extraction was carried out by combining NDVI with land surface temperature, based on UAV multispectral and thermal infrared

data. Pixels with NDVI values below 0.1 were classified as non-vegetation, while those with values above 0.1 were classified as vegetation [45]. Based on in situ measurements of soil and mulching-film temperatures and empirical observations distinguishing cropland from non-cropland, we set the farmland temperature thresholds to 5–15 °C in April, 10–20 °C in May, and 15–25 °C in June. As the key distinction between natural vegetation and crops lies in their canopy heights, UAV LiDAR data were employed to extract the Canopy Height Model (CHM) [46]. This was used to further differentiate farmland from natural vegetation. Specifically, the CHM feature was computed as the elevation difference (DSM − DTM), with both DSM and bare-earth DTM derived from LiDAR data. During data acquisition, three-return recording was enabled, with an average point-cloud density of 540 points/m^2,and we used Triangular Irregular Network (TIN)-based linear interpolation for the DSM and DTM [47]. Based on the temporal patterns of crops in the Demonstration Zone and measured plant heights, farmland was selected according to the following CHM criteria: ≤10 cm in April (with an additional threshold of 65–85 cm for identifying winter wheat), ≤60 cm in May, and ≤120 cm in June. Subsequently, the SAM and MRS methods were employed to segment the extracted RGB farmland data into plots, thereby facilitating early plot identification. SAM performed full-automatic segmentation of the entire image, and key parameters were adjusted (Section 3.3) to optimize the segmentation results. To evaluate extraction accuracy, ground-marked farmland points on the image were used as reference annotations. A confusion matrix was then applied to calculate farmland extraction accuracy, false alarm rate, and overall performance. Farmland extraction accuracy represents the probability of correctly identifying farmland, reflecting the model's detection capability. The false alarm rate indicates the proportion of non-farmland misclassified as farmland, while overall performance accuracy quantifies the proportion of correctly classified instances, representing the model's overall performance [24].

3.2.2. Optimal Feature Selection

Random Forest [48] is a powerful machine learning algorithm whose built-in feature importance assessment aids in identifying key features and reducing redundancy from excessive spectral and texture information [49]. It is inherently robust and resistant to noise, as feature importance is based on the mean vote of multiple trees.

In this study, the process for optimal feature selection was as follows: for each tree T, the out-of-bag (OOB) error was calculated and denoted as $Error_T^{original}$. Next, the values of feature j within the OOB samples were randomly permuted, and the error $Error_T^{permuted(j)}$ was recalculated. Feature importance was defined as the mean increase in error after permutation. The importance *PIj* was calculated as:

$$PI_j = \frac{1}{N_T} \sum_{T=1}^{N_T} \left(Error_T^{permuted\ (j)} - Error\ _T^{original} \right) \quad (1)$$

At the ground sample statistics stage, 75 ground points were recorded. Based on these point locations, 200 sample points were extracted by interpreting spectral and texture features from six UAV data periods. The RF algorithm was then applied to identify the most important features. Of the sample points, 80% were used for training and 20% for validation. Eleven spectral features were calculated using corresponding band functions, with correction parameters set at 0.5 for crop vegetation. The six-band multispectral data generated eight texture features per band. Thus, each data period produced 59 features for optimal feature selection. From these, the ten most correlated features per period were selected, and their summed correlation proportions were compared. The optimally selected features were then used for vegetation structure extraction, with a planned computational load of ten features. For image data with insufficient importance during the

initial selection, comparative operations between datasets were conducted, followed by another round of feature selection. By amplifying characteristic differences among planting structures, optimal feature selection was achieved, enabling earlier extraction of planting structure data.

3.2.3. Planting Structure Extraction

RF performs favorably in high-dimensional feature spaces, whereas support vector machines (SVMs) generally exhibit stronger generalization under small-sample conditions. Because these properties align well with the characteristics of UAV-derived data, we employ RF and SVM, together with a deep-learning convolutional neural network (CNN), to conduct crop-structure mapping using optimally selected features. For each method, 200 visually interpreted sample points were collected, with 160 used for training and 40 for validation.

(1) Random Forest (RF)

RF is an ensemble learning algorithm that classifies data by constructing multiple decision trees and aggregating their predictions. The final classification result is obtained through majority voting or probability averaging [48]:

$$P(y \mid x) = \frac{1}{T}\sum_{t=1}^{T} P_t(y \mid x) \tag{2}$$

where $P(y \mid x)$ represents the probability that input x belongs to class y, estimated by the t-th tree. In this study, RF classification relied on two core mechanisms. The first was a bootstrap sampling strategy that generated multiple heterogeneous subsets from the original training set through sampling with replacement, enhancing classifier diversity. During individual tree construction, the optimal splitting rule was determined from a randomly selected feature subset (typically $\sqrt{\mathrm{M}}$, where M is the number of features), reducing correlation among trees.

The second mechanism was ensemble prediction via majority voting, where final predictions were obtained by aggregating all base classifiers. This approach statistically minimizes expected classifier errors, reducing overfitting risk.

(2) SVM

By applying the kernel trick, SVM mapped selected textural and spectral features into a high-dimensional space [50]. The optimization problem of constructing a maximum-margin hyperplane was transformed into its dual form:

$$\max_{\alpha} \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \; s.t.\; C \geq \alpha_i \geq 0, \sum_{i=1}^{m} \alpha_i y_i = 0 \tag{3}$$

where m denotes the number of training samples; x_i is the feature vector of the i-th sample; $y_i \in \{-1, +1\}$ is the class label; α_i is the Lagrange multiplier, $C > 0$ is the penalty parameter and $K(x_i, x_j)$ is the kernel function. The decision function was:

$$h(x) = \mathrm{sgn}\left(\sum_{j \in S} \alpha_j y_j K(x_j, x) + b\right) \tag{4}$$

where $b \in R$ denotes the bias term. S is the set of support vectors (i.e., samples with $\alpha_j \neq 0$):

$$\begin{gathered} \text{Let } S = \{j \mid \alpha_j \neq 0\} \text{ be the set of support vectors. Then} \\ h(\mathrm{x}) = \sum_{j \in S} \alpha_j y_j K(\mathrm{x}_j, \mathrm{x}) \end{gathered} \tag{5}$$

(3) Convolutional Neural Network (CNN)

The CNN consists of convolutional, pooling, activation, and fully connected layers. Its core principle is to reduce parameter scale through local connections and weight sharing while extracting image features [51]. In this study, CNN progressively extracted high-level semantic features of ground objects from vegetation indices and texture features through stacked convolutional operations.

For the input image I and convolution kernel K in the convolutional layer, the output feature map O was calculated using the following formula:

$$O(i,j) = \sum_{m=0}^{k_h-1} \sum_{n=0}^{k_{tw}-1} I(i+m, j+n) \cdot K(m,n) + b \quad (6)$$

where $k_h \times k_w$ is the size of the convolution kernel, and b is the bias term. For multispectral input, the kernel was extended into a 3D tensor to compute channel-weighted sums.

In the fully connected layer, 2D convolutional features were flattened into a 1D vector and mapped into class space via the Softmax function:

$$P(y=c) = \frac{e^{z_c}}{\sum_{k=1}^{C} e^{z_k}}, z = W \cdot x + b \quad (7)$$

where z_c represents the class score, and W represents the weight of the fully connected layer.

3.2.4. Accuracy Assessment

In the accuracy assessment stage, samples were divided into training and validation sets through stratified random sampling. Validation sample vectors were overlaid onto classification results to generate a confusion matrix and compute accuracy indices. These accuracy indices included Overall Accuracy (OA), Kappa Coefficient [52], as well as class-level Producer's Accuracy (PA) and User's Accuracy (UA). OA comprehensively reflects the proportion of correctly classified pixels. The Kappa Coefficient measures the consistency between the classification results and the ground truth data. PA and UA indicate the classification performance for individual classes from the perspectives of errors caused by classification omission (producer's perspective) and false classification (user's perspective), respectively. The formulas for calculating these indices are given below:

$$OA = \frac{\sum_{i=1}^{n} TP_i}{N} \times 100\% \quad (8)$$

$$Kappa = \frac{N\sum_{i=1}^{n} X_{ii} - \sum_{i=1}^{n}(X_{i+} + X_{+i})}{N^2 - \sum_{i=1}^{n}(X_{i+} + X_{+i})} \quad (9)$$

$$PA = \frac{X_{jj}}{X_{+j}} \quad (10)$$

$$UA = \frac{X_{ij}}{X_{i+}} \quad (11)$$

where N represents the total number of samples, TPi represents the number of correctly classified samples for the i-th class, and n represents the number of classes.

Significance testing was conducted for the classification results across different methods. Overall Accuracy (OA) was pre-specified as the primary comparison metric. Pairwise comparisons were performed on the same set of objects: a stratified paired bootstrap with B = 10,000 resamples (stratified by crop class, at the object level) was used to estimate the 95% confidence interval for ΔOA, and a paired permutation test with M = 50,000 permutations was used to obtain two-sided *p*-values for ΔOA. Because only a single pre-specified primary

comparison (OA) was considered, no multiple-comparison correction was applied; the significance level was set at $\alpha = 0.05$.

3.3. Experimental Setting

For feature extraction (RF for feature importance/selection),we set n_estimators = 100 (to balance stability and runtime), max_depth = None (to allow the trees to grow and capture nonlinearity), max_features = 8 (following the $\sqrt{p}$ heuristic; set to 8 in our data to reduce variance and improve generalization), and min_samples_split = 2 (to limit excessive splits and suppress noise). These choices were informed by high-performing prior studies and our preliminary trials [53,54]. All other parameters used default values.

As for classification, the input for each classification consists of the top-10 ranked features. For RF, we used n_estimators = 500, max_depth = None, max_features = 8, and min_samples_split = 5. For SVM, we set C = 10 (a stronger penalty suitable for high-resolution imagery) and $\gamma = 0.01$ (a smaller kernel coefficient for multi-dimensional features; RBF kernel) [55,56]. All remaining parameters were kept at their default settings.

In the field of CNN, for classification, the ten most informative features are retained. With eight planting-structure classes, each sample is represented as an 8×10 input matrix. The network architecture consists of two convolutional layers (3×3 kernels, 8 channels each), followed by a max-pooling layer and a flattening operation that yields a 72-unit feature vector. This vector is fed to a fully connected layer with 32 units and then to a final dense output layer with a softmax activation to produce class probabilities over the eight categories. Model parameters are optimized using the Adam optimizer [57].

Table 3 lists the key parameters of the SAM segmentation model and their roles. points_per_side determines the grid sampling density; increasing it appropriately can improve the integrity of small objects and fine boundaries, at the cost of higher computation. crop_n_layers enhances recall in dense regions through hierarchical cropping, which is beneficial for complex textures and mixed plantings, but excessive layers may introduce redundancy and over-fragmentation. pred_iou_thresh and stability_score_thresh jointly serve as a quality–stability gate for candidate masks: values that are too low may introduce noise, whereas values that are too high can suppress the recall of fine fragments. crop_n_points_downscale_factor primarily affects inference efficiency; a value of 2 is generally more robust provided it does not cause noticeable quality fluctuations. min_mask_region_area suppresses spurious small patches and holes, but an overly large threshold may erroneously remove true small objects [24,58,59].

Table 3. Key hyperparameters of the SAM model.

Parameters Name	Final Parameters	Explanation
points_per_side	128	Defines sampling points along one image side.
pred_iou_thresh	0.86	Filters masks based on predicted quality, within the range [0, 1].
stability_score_thresh	0.92	Adjusts cutoff for stability score calculation.
crop_n_layers	1	Determines the number of image crop layers, where each layer includes $2i$ image subdivisions.
crop_n_points_downscale_factor	2	Downscales sampled points per side for layer n by a factor of $2n$.
min_mask_region_area	80	Removes small regions and holes in masks smaller than the specified area.

The sensitivity analysis was conducted using the same data and implementation environment as the main experiments. Candidate parameters were first selected based on segmentation-accuracy evaluations, after which the search ranges and step settings were

as follows:points_per_side ∈ {96, 128, 160, 192, 224}; pred_iou_thresh ∈ {0.80, 0.85, 0.90, 0.92, 0.95}; stability_score_thresh ∈ {0.88, 0.90, 0.92, 0.94, 0.96}; crop_n_layers ∈ {0, 1, 2, 3, 4}; crop_n_points_downscale_factor ∈ {1, 2, 3, 4, 5}; min_mask_region_area ∈ {40, 60, 80, 100, 120} (px^2). Evaluation was performed on the June data, which provided the best segmentation performance. Repeated experiments indicate the optimal configuration is: points_per_side = 128–160, pred_iou_thresh = 0.86–0.90, stability_score_thresh = 0.90–0.94, crop_n_layers = 1–2, crop_n_points_downscale_factor = 2, and min_mask_region_area = 80.

Multiresolution segmentation (MRS) is a commonly used object-oriented segmentation approach. By tuning three parameters—scale, shape, and compactness—optimal results can be achieved [60]. Based on multiple experiments, the optimal parameter values were determined to be 200, 0.1, and 0.5, respectively.

The SAM image segmentation was implemented using the PyTorch framework. The proposed method was executed in an environment with Python 3.10.12, PyTorch 2.2.0, and CUDA 11.7, and all experiments—including the sensitivity analysis—were conducted on an NVIDIA Tesla A40 GPU.

4. Results and Analysis

4.1. Farmland and Plot Extraction Result

4.1.1. Farmland Extraction Results

UAV multispectral and LiDAR data from April, May, and June were compared. The results showed that the combination of thermal infrared (TIR) data and vegetation indices (NDVI) could consistently distinguish between "vegetation" and "non-vegetation" land cover types (Figure 3). However, the accuracy of farmland extraction was limited in the early planting stage due to the use of mulching film. After integrating the CHM, it became possible to further differentiate "natural vegetation" from "farmland crops" within the vegetation class. This integration significantly improved discrimination accuracy between mulching film and non-farmland areas, as well as between natural vegetation and farmland. Under the "TIR + NDVI + CHM" configuration, the OA and Kappa coefficient were 82.1% and 0.732, respectively, in April; increased to 89.3% and 0.851 in May; and further improved to 96.4% and 0.946 in June. Month by month, OA increased by 7.2% from April to May and by another 7.1% from May to June, with a cumulative increase of 14.3% from April to June. The Kappa coefficient rose from 0.732 to 0.851 (an increase of 0.119) and further to 0.946 in June (an increase of 0.214 compared to April) (Table 4). With the inclusion of CHM, the configuration already showed preliminary farmland extraction capability in April and could be applied for rapid early surveys. By May, the data achieved high reliability, enabling high-precision statistical results. June was determined as the optimal phenological period, making it the most suitable reference period for producing base maps and fine-scale farmland maps. Within the same period, farmland extraction accuracy improved significantly after coupling "TIR + NDVI" with CHM, confirming that the incorporation of vertical structural information into UAV multi-source data fusion was essential for improving early farmland identification.

Table 4. Accuracy of parcel delineation from multi-sensor UAV data across April, May, and June.

Accuracy Statistics	NDVI	TIR	NDVI + TIR	NDVI + TIR + CHM
April	0.531	0.359	0.562	0.821
May	0.575	0.492	0.591	0.893
June	0.665	0.593	0.706	0.964

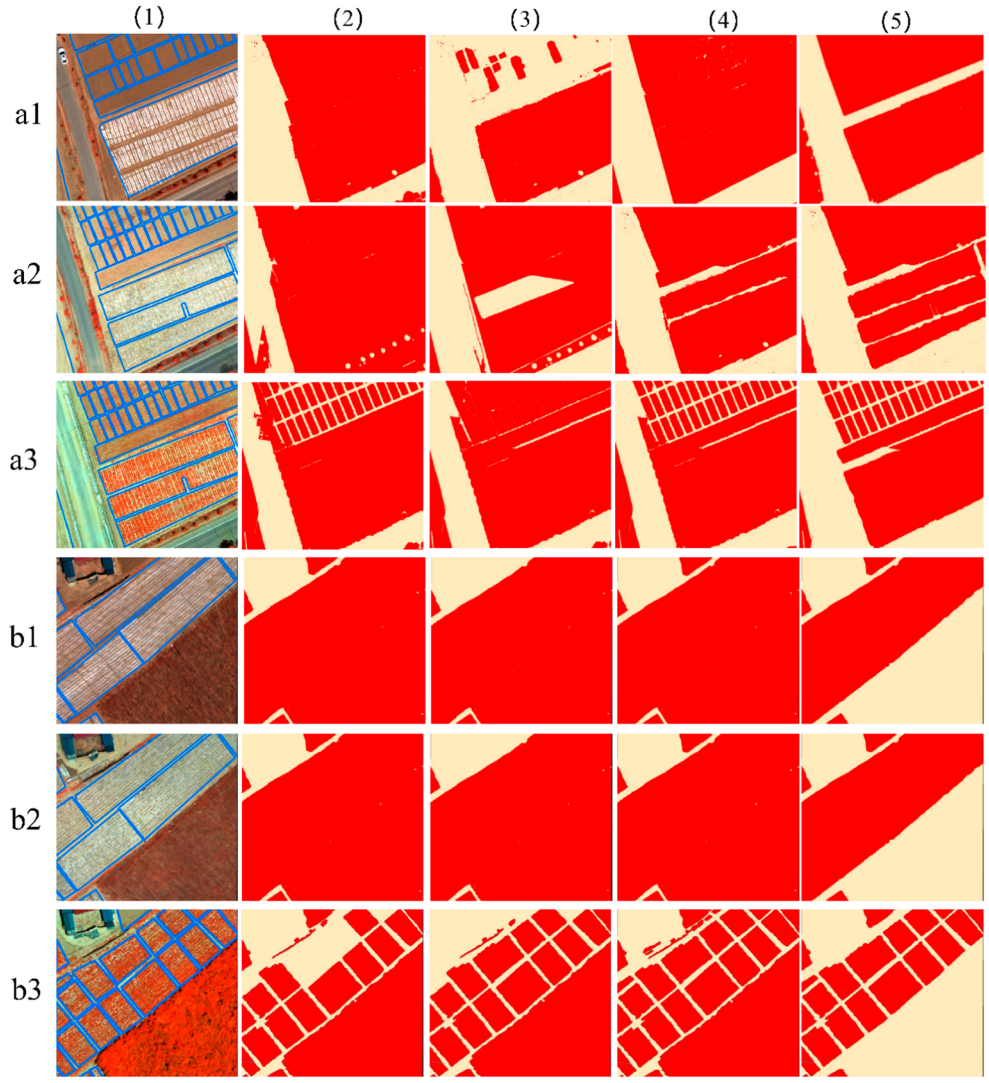


Figure 3. Single-period cropland extraction results: (**a1**–**a3**), Parcel L1 (Apr–Jun); (**b1**–**b3**), Parcel L2 (Apr–Jun). (**1**) False-color image; (**2**) NDVI-based; (**3**) TIR-based; (**4**) NDVI + TIR-based; (**5**) NDVI + TIR + CHM-based.

4.1.2. Plot Extraction Result

The identified farmland was further subdivided into monoculture objects, and detection accuracy (Detection), false alarm rate (FAR), and pixel-level overall accuracy (OA) were calculated. A comparison of April–June data revealed that SAM achieved the best segmentation performance in June, with Detection = 0.874, FAR = 0.094, and OA = 0.928 (Table 5). This performance was markedly superior to MRS, which yielded Detection = 0.589, FAR = 0.403, and OA = 0.627. To enhance visual discernibility of plot boundaries, false color images from July were used as the background (since some plots were still covered with mulching film in June) (Figure 4). Based on these results, June was confirmed as the optimal phenological period for object segmentation. The plot units generated in June supported the precise extraction and statistical mapping of early-stage planting structures.

Table 5. Segmentation accuracy of SAM in April–June.

Accuracy Statistics	SAM-June	SAM-May	SAM-April
Detection accuracy	0.874	0.721	0.683
False alarm rate	0.094	0.158	0.209
Overall accuracy	0.928	0.832	0.791

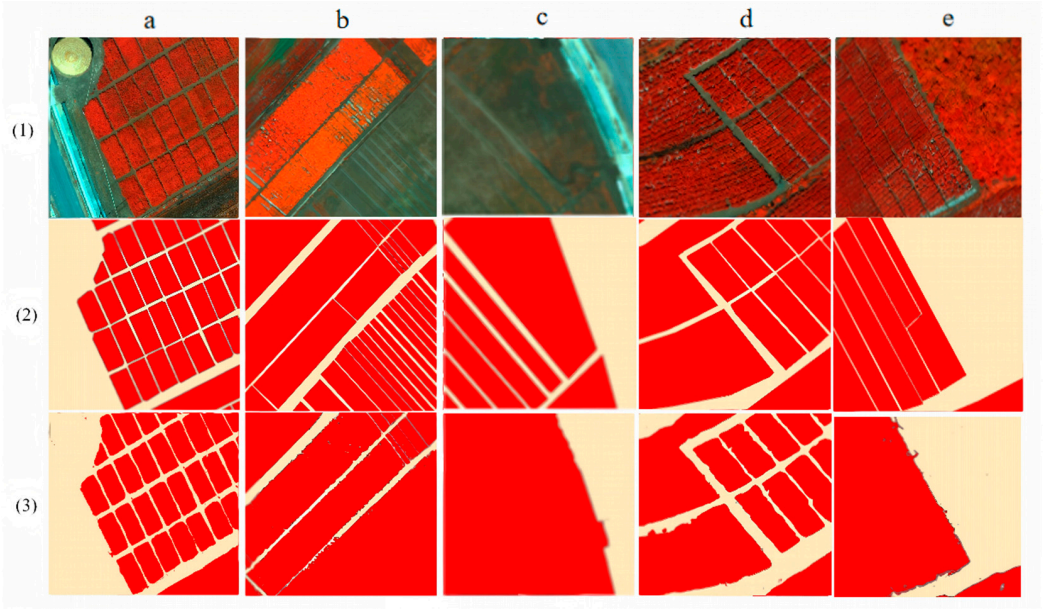


Figure 4. Parcel segmentation results. Columns (**a**–**e**) denote five representative plots spanning different sizes and shapes: (**1**) July false-color composites; (**2**) ground-truth delineations; (**3**) June SAM-based segmentation outputs.

4.2. Optimal Feature Selection Results

4.2.1. Optimally Selected Features for the Entire Planting Structure

Figure 5 and Table 6 present the optimal features selected by the RF algorithm across six image periods. The cumulative importance weight of the top ten features peaked in July at 72.26%. For the remaining periods, the sums of weights decreased in the following order: 71.07% in August, 70.83% in September, 51.95% in June, 21.95% in May, and 16.39% in April.

To enable earlier identification of planting structures, optimal feature selection was performed by comparing data from periods prior to July. The cumulative importance of the top ten features selected from April and June data was 67.36%, only 4.9% lower than in July. In contrast, the cumulative importance for May and June was 58.52%, and for April and May, it was 43.27%, both showing substantial differences from July's results.

4.2.2. Optimally Selected Features for Monoculture Structures

For monoculture structures, optimal features were selected monthly from UAV data to determine key features for early planting structure identification (Figure 6). July was identified as the optimal period for selecting features for maize, maize-legume intercropping, millet, legumes, vegetables, and sorghum. During this period, the importance of the top ten features for these crops accounted for 82%, 86%, 79%, 85%, 83%, and 80%, respectively. Compared with the ratio features selected from April and June, the importance weights increased by 8% (millet), 19% (legumes), 3% (vegetables), 9% (sorghum), 4% (maize), and 13% (maize-legume intercropping) (Table 7). Thus, apart from legumes and maize-legume intercropping, the planting structures of the remaining four crops could also be identified early in June using the ratio features from April and June. Before July, buckwheat had not been sown; its optimal features were identified in September, with the top ten features accounting for 87% of importance, an 8% increase compared with August. Wheat was harvested in May, and this month was identified as the optimal period for its feature selection, with the top ten features accounting for 81% of importance, only 8% higher than in April. Therefore, wheat planting structures could also be extracted rapidly in April.

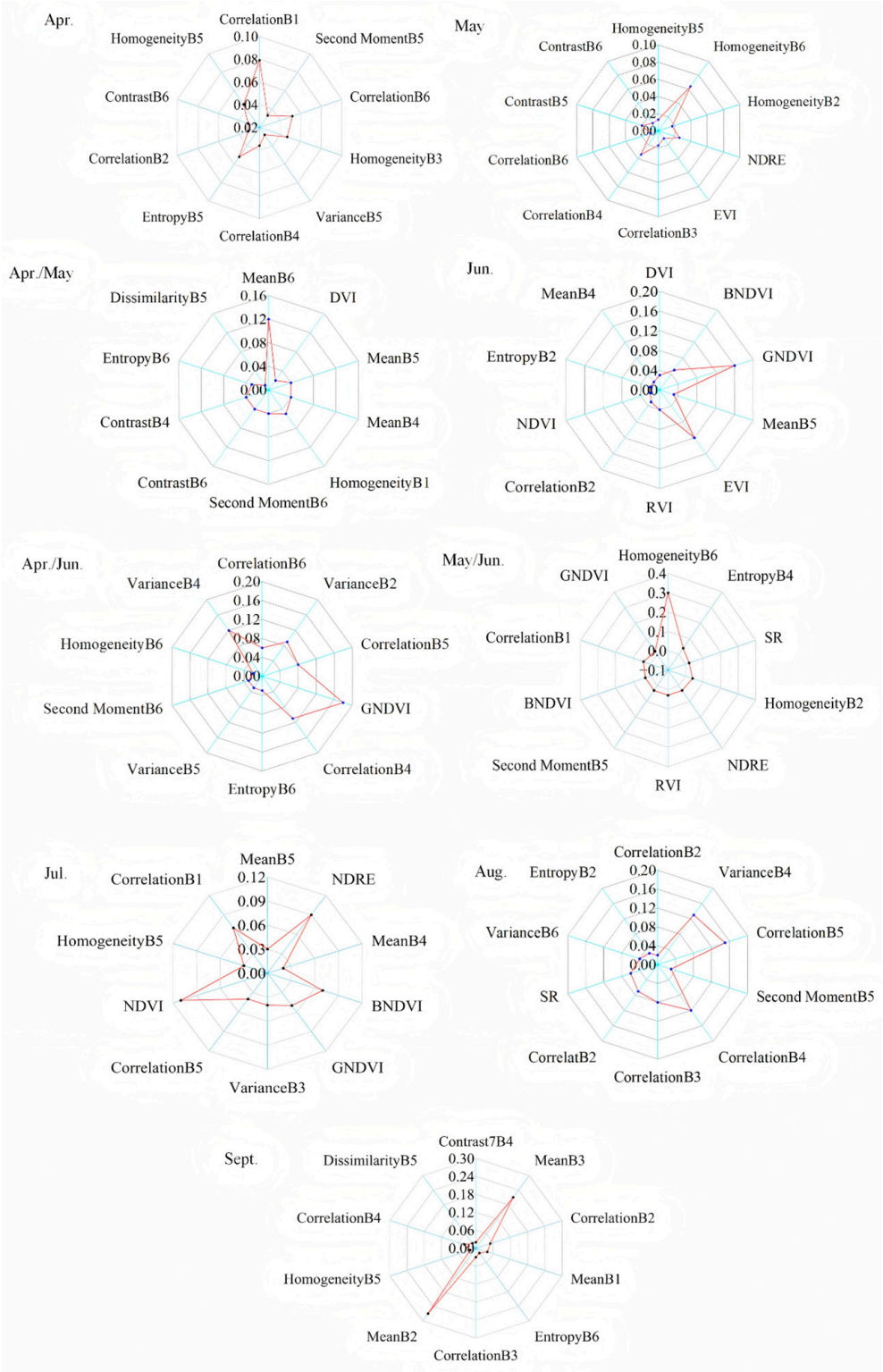


Figure 5. The Entire Planting Structure: Top-10 Correlation-Ranked Features across April–September and across Pairwise Comparisons of Pre-July Months.

Table 6. Correlation-based share of the top-10 features for the entire planting structure (Apr–Sep).

Period	April	May	Apr./May	June	Apr./Jun.	May/Jun.	July	August	September
Sum of Top10	16.39%	21.95%	43.27%	51.95%	67.36%	58.52%	72.26%	71.07%	70.83%

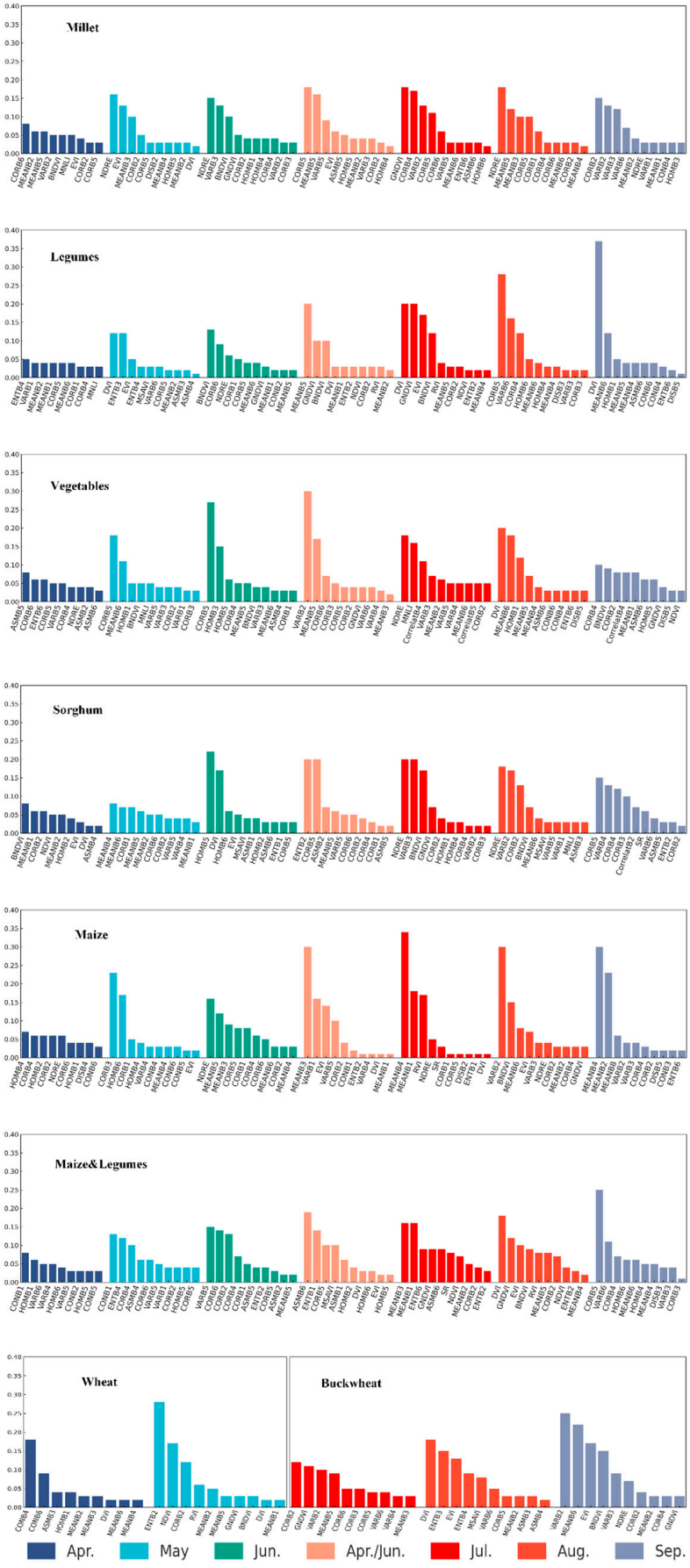


Figure 6. Monthly top-10 feature importances for monoculture structures.

Table 7. Correlation-based proportion of the top-10 features for monoculture structures (April–September).

Crop\Period	April	May	June	Apr./Jun.	July	August	September
Millet	0.58	0.61	0.65	0.71	0.79	0.7	0.66
Legumes	0.39	0.45	0.5	0.66	0.85	0.77	0.76
Vegetables	0.54	0.62	0.75	0.80	0.83	0.76	0.65
Sorghum	0.5	0.53	0.7	0.71	0.8	0.74	0.75
Maize	0.57	0.65	0.73	0.78	0.82	0.8	0.78
Maize & Legumes	0.49	0.68	0.69	0.73	0.86	0.81	0.74
Wheat	0.73	0.81	-	-	-	-	-
Buckwheat	-	-	-	-	0.66	0.79	0.87

4.3. Classification Results and Accuracy Assessment

Using ground-marked point locations, the crop planting structures were classified into nine categories: legumes, millet, maize, buckwheat, wheat, sorghum, maize-legume striped intercropping, vegetables, and others. Among these, the SAM + RF method achieved the highest accuracy for planting structure extraction in July, with OA = 92.66% and Kappa = 0.9143. This accuracy was positively correlated with the weights of the optimally selected features (Figure 7).

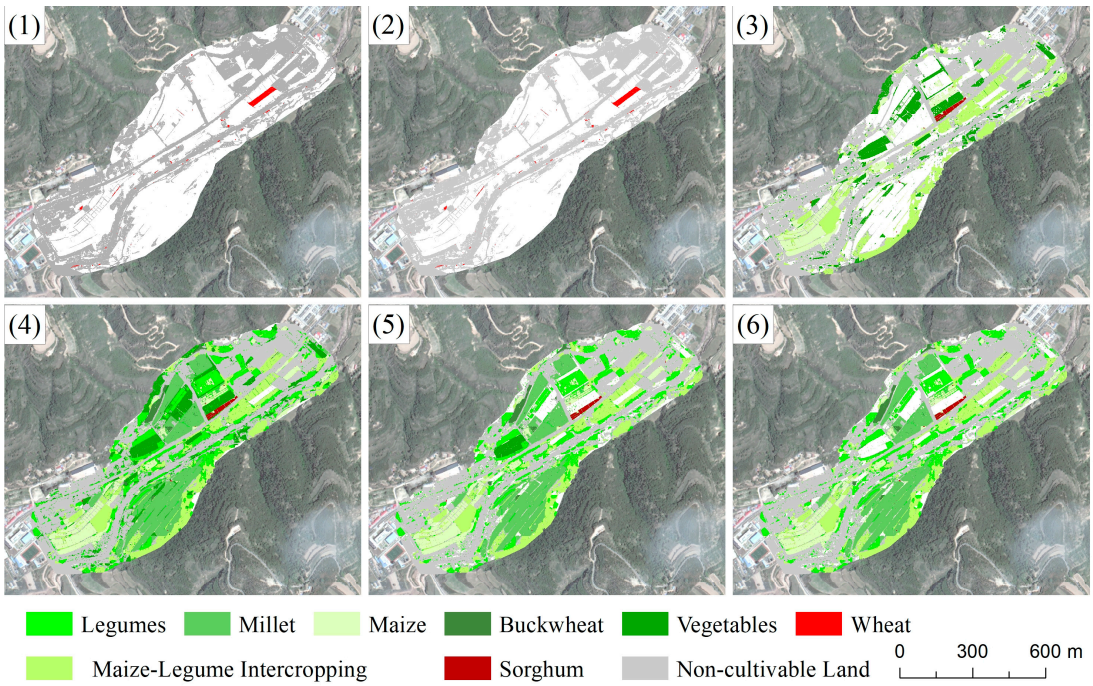


Figure 7. Panels (**1**–**6**) present classification results for the entire planting structure from April to September.

For monoculture structures, the highest classification accuracy for wheat was achieved in May (98%), with 5.12 mu (0.34 ha) (mu, Chinese acre, 1 mu = 0.067 ha) of planting area identified; in April, 4.95 mu (0.33 ha) of wheat planting area was identified with 94% accuracy. For buckwheat, classification accuracy reached 98% in September, with 13.89 mu (0.93 ha) of planting area identified; in August, 12.15 mu (0.81 ha) were identified with 86% accuracy. For other crops, July provided the highest accuracies: 114.17 mu (7.61 ha) of millet (93%), 105.36 mu (7.02 ha) of legumes (97%), 79.89 mu (5.33 ha) of vegetables (97%), 3.44 mu (0.23 ha) of sorghum (97%), 53.89 mu (3.59 ha) of maize (96%), and 105.21 mu (7.01 ha) of maize-legume intercropping (96%) (Figure 8 and Table 8).

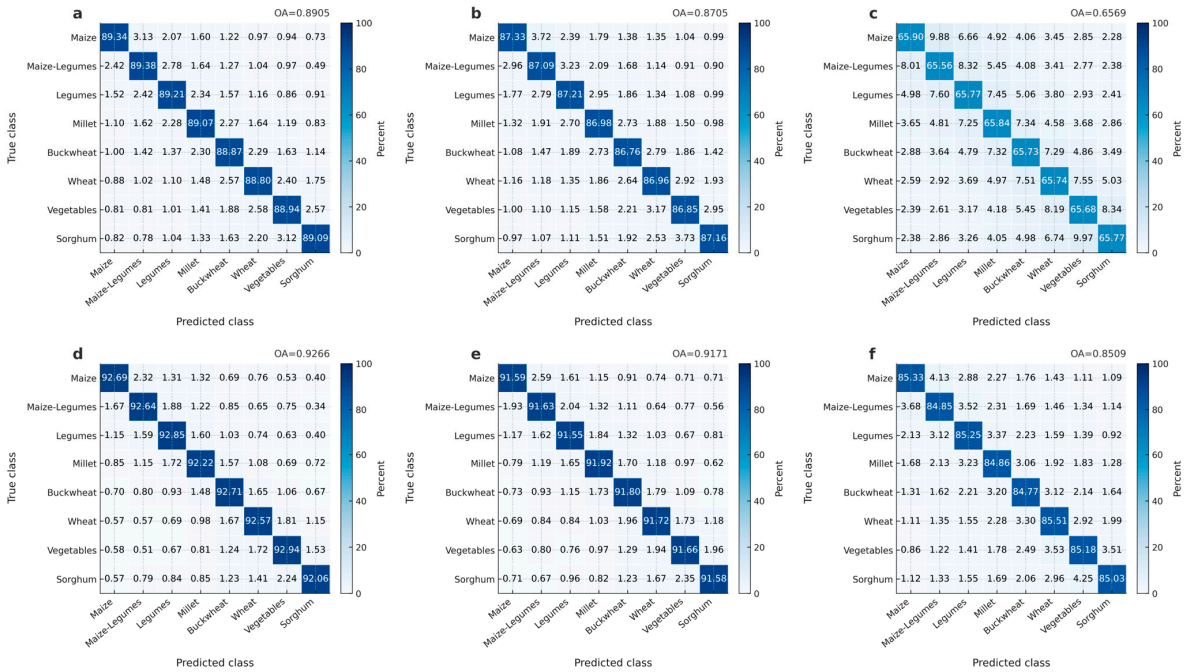


Figure 8. Confusion matrices for July classifications using MRS- and SAM-based segmentations combined with RF, SVM, and CNN classifiers: (**a**) MRS + RF, (**b**) MRS + SVM, (**c**) MRS + CNN, (**d**) SAM + RF, (**e**) SAM + SVM, (**f**) SAM + CNN.

Table 8. Area statistics, classification accuracy, and NDVI for monoculture structures (April–September). (a) Classification Area (ha) for Monoculture Structures (April–September). (b) Classification Accuracy for Monoculture Structures (April–September). (c) NDVI for Monoculture Structures (April–September).

(a)

	April	May	June	July	August	September
Millet	0.011	0.214	3.743	7.611	6.810	6.393
Legumes	0.023	1.957	2.881	7.024	6.759	6.597
Vegetables	0.142	1.519	3.346	5.326	5.065	4.821
Sorghum	0.006	0.086	0.145	0.229	0.207	0.177
Maize	0.047	1.065	1.909	3.593	3.249	3.087
Maize & Legumes	0.010	2.053	3.960	7.014	6.597	6.145
Wheat	0.330	0.341				
Buckwheat				0.688	0.810	0.926

(b)

	April	May	June	July	August	September
Millet	0	0.03	0.46	0.93	0.83	0.78
Legumes	0	0.27	0.4	0.97	0.94	0.91
Vegetables	0.03	0.28	0.61	0.97	0.92	0.88
Sorghum	0.03	0.36	0.61	0.97	0.87	0.74
Maize	0.01	0.28	0.51	0.96	0.87	0.82
Maize & Legumes	0	0.28	0.54	0.96	0.9	0.84
Wheat	0.94	0.98				
Buckwheat				0.73	0.86	0.98

(c)

	April	May	June	July	August	September
Millet	0.02	0.15	0.27	0.55	0.62	0.35
Legumes	0.01	0.1	0.18	0.22	0.88	0.6
Vegetables	−0.1	0.1	0.56	0.82	0.03	0
Sorghum	0.04	0.1	0.35	0.78	0.7	0.2
Maize	0.05	0.15	0.65	0.85	0.75	0.35
Maize & Legumes	0.02	0.12	0.52	0.78	0.5	0.1
Wheat	0.75					
Buckwheat				0.76	0.78	0.82

Comparative experiments were also conducted using July and August data, applying SAM and MRS segmentation methods combined with RF, SVM, and CNN. The results showed that (1) SAM-based classifications consistently outperformed MRS-based classifications and (2) July images produced higher accuracy than August images, regardless of method. With the same segmentation method, RF consistently outperformed SVM and CNN in classification accuracy. Using the same data partition and the same 10-feature best subset, we trained SAM/MRS + RF/SVM/CNN for July and August and computed validation-set OA and Kappa; results appear in Table 9 (a). According to the significance testing results, for July, compared with the best baseline, SAM + RF achieved ΔOA = +0.95% (95% CI: +0.10%–+1.80%), with $p = 0.036$ from the paired permutation test. For August, compared with the best baseline, SAM + RF achieved ΔOA = +4.14% (95% CI: +2.80%–+5.50%), with $p < 0.001$ from the paired permutation test (Table 9 (b)).

Table 9. (a). July–August classification accuracies for MRS- and SAM-based segmentations combined with RF, SVM, and CNN classifiers. (b) Significance tests for July–August classification-accuracy comparisons of RF, SVM, and CNN built on MRS- and SAM-based segmentations.

(a)

Segmentation Methods	Classification Methods	Kappa Coefficient		Overall Accuracy	
		July	August	July	August
MRS	RF	0.8919	0.8276	0.8905	0.8351
	SVM	0.8691	0.7843	0.8705	0.7903
	CNN	0.6356	0.5917	0.6569	0.6124
SAM	RF	0.9163	0.8718	0.9266	0.8847
	SVM	0.9051	0.8505	0.9171	0.8433
	CNN	0.8293	0.7817	0.8509	0.7942

(b)

Month	Comparator Method	OA (SAM + RF)	OA (Comparator)	ΔOA = OA_SAM + RF − OA_Comp.	95% CI for ΔOA (%)	*p* (Paired Permutation)
Jul	Best baseline (SAM + SVM)	0.9266	0.9171	0.0095	+0.10–+1.80	0.036
Jul	MRS + RF	0.9266	0.8905	0.0361	+2.20–+4.95	<0.001
Jul	SAM + SVM	0.9266	0.9171	0.0095	+0.10–+1.80	0.036
Jul	SAM + CNN	0.9266	0.8509	0.0757	+6.20–+8.90	<0.001
Aug	Best baseline (SAM + SVM)	0.8847	0.8433	0.0414	+2.80–+5.50	<0.001
Aug	MRS + RF	0.8847	0.8351	0.0496	+3.60–+6.30	<0.001
Aug	SAM + SVM	0.8847	0.8433	0.0414	+2.80–+5.50	<0.001
Aug	SAM + CNN	0.8847	0.7942	0.0905	+7.80–+10.30	<0.001

The statistical area, classification accuracy, and NDVI for monoculture structures were calculated, and the results are presented in Table 8.

5. Discussion

5.1. Crop Phenology and Early Identification of Planting Structures

From the perspective of classification accuracy, wheat was harvested in May, but its classification accuracy in April had already reached 94%, indicating that wheat could be precisely identified during early farmland extraction. Buckwheat was sown in July, and its optimally selected features from the subsequent three periods showed a rising trend. Specifically, buckwheat identification accuracy reached 73% in July, enabling preliminary identification; in August, accuracy improved by 13%, entering the stage of precise identification. Most other crops were in their early growth stages in April and May and reached maturity by July. After July, these crops entered harvesting or decay stages, during

which both optimally selected features and vegetation indices showed a declining trend. Thus, July was determined as the key period for high-precision identification of planting structures in the Demonstration Zone (Table 8). Additionally, the features optimally selected for maize, legumes, and maize-legume intercropping exhibited large differences within the same period. In April and May, maize and legumes had weaker feature sets compared with maize-legume intercropping, for which texture features dominated. This indicated that texture features played a positive role in early identification of maize-legume intercropping during the mulching film stage. These results indicate that the strong and variable background and specular-reflection signals introduced by plastic mulch mask crop canopy spectral differences, leading to a marked decline in canopy spectral separability [61]. Overall, texture features dominated before June across all monoculture structures. After June, as crops entered growth and maturity stages, the proportion of vegetation indices among optimally selected features increased significantly. This reflected that the morphological structure of growing crops was more sensitive to environmental responses and exhibited greater variability, whereas mature crops were more closely associated with vegetation indices [62]. However, based on the overall planting-structure feature-importance analysis, the contribution of vegetation indices (VIs) was lowest in April (0%), increased to 18.33% in May, rose to 66.86% in June, and then declined to 47.5% in July. This indicates that VIs approach saturation in July, inter-crop spectral contrasts diminish, and structural differences among crops emerge—appearing as more pronounced texture cues in high-resolution imagery.

For the entire planting structure, both the optimal feature selection results and classification accuracy in July were outstanding, mainly due to two factors: (1) Significant crop phenological differences. In July, major crops such as maize, legumes, maize-legume intercropping, sorghum, millet, and vegetables were mature, while buckwheat was newly sown and sparse weeds appeared after wheat harvest. At this stage, vegetation indices varied substantially among planting structures, and texture features exhibited enhanced directionality and differences in canopy densities. Consequently, vegetation indices along with texture Mean and Correlation values were highly important among RF-selected features. (2) High vegetation coverage. In July, major crops exhibited high LAI. Correspondingly, NDVI and NDRE were identified as key vegetation indices. The NDVI plays an important role in quantifying canopy biophysical parameters (chlorophyll and water content) and is strongly correlated with plant physiological status (e.g., photosynthetic efficiency) [63]. Accordingly, its prominence as the most relevant feature for July classification across all planting structures reflects marked between-crop differences in chlorophyll and water content, as well as in photosynthetic rates, during that month. By contrast, the NDRE is more effective than NDVI for chlorophyll retrieval under high-biomass conditions and is less prone to saturation [64]; moreover, its responsiveness to water stress is markedly higher than that of NDVI in the UAV context [65]. Hence, the fact that NDRE ranked as the second most relevant feature further indicates that at this stage—when fields have entered a relatively high-biomass state—substantial inter-crop differences exist in both water status and chlorophyll content. From September onwards, crops had entered specific growth stages, and the differences in spectral characteristics between crops decreased. Hence, texture features became dominant among optimally selected features. This result further confirmed the effectiveness of multi-temporal data fusion of spectral and texture features in distinguishing planting structures. Previous studies reported that vegetation indices are prioritized over texture features when low-resolution, large-scale remote sensing data are applied [66,67]. However, in this study, the small research area and high-resolution UAV data resulted in texture features contributing more significantly than vegetation indices across all stages. This suggests that texture features become less distinguishable

at low resolutions or in larger research areas, whereas significant differences can be observed when high-resolution UAV data are used [68]. It contends that UAV imagery (typically centimeter to sub-meter resolution) can clearly capture fine surface structures (e.g., leaf arrangement, soil granules), which provide effective textural cues. In contrast, at medium-resolution satellite scales (e.g., 10 m), individual pixels mix multiple land-cover spectra, blurring texture and erasing heterogeneity, thereby reducing the contribution of texture to classification [69]. Accordingly, given the pronounced resolution disparities among heterogeneous multi-source remote-sensing data, while we consider reusing the scheme developed in this study—the UAV-derived feature set and extraction time window—at broader satellite scales, the choice of cross-scale fusion methods must be carefully addressed [11]. If the adopted transformation algorithms [70] or spatio-temporal resampling/interpolation approaches [71] can effectively leverage UAV data to enhance the accuracy of vegetation-parameter retrieval from satellite imagery, the features identified here will acquire substantially greater reference value for large-area applications.

5.2. Timeline and Feature Selection for Early Identification of Planting Structures

Through optimal feature selection and result analysis, a fine-scale identification timeline for typical planting structures on the Loess Plateau based on UAV data was established: farmland and winter wheat could be identified in April, farmland plots in June, and buckwheat, maize, legumes, maize-legume intercropping, vegetables, millet, and sorghum in July. Between June and July, most crops underwent rapid growth. Considering the low cost, high maneuverability, and light computational load of UAV platforms, a 10-day high-time-resolution multi-source data collection strategy is recommended during this period to capture the critical timeline for early crop identification. According to the optimally selected features for the entire planting structure, crops in June were in the early growth stage and exhibited remarkable spectral differences. Planting structure extraction at this stage relied mainly on vegetation indices (GNDVI, NDVI) and texture features such as Mean. By July, crops had matured, and while the overall composition of optimally selected features remained similar, the importance of Cor features increased significantly. Cor features assess the correlation between gray levels of neighboring pixels, reflecting local texture consistency [42]. Their increased importance in July was directly related to the high uniformity of crop canopy textures during maturity, confirming a strong relationship between Cor features and crop maturity. Hence, UAV-derived Cor data could serve as an indicator for predicting crop maturity [72]. Notably, the difference between features optimally selected from April and June and those from July was less than 5%. This indicated that rapid feature ratio analysis using April and June multispectral data could be employed for timely but less precise planting structure identification if high timeliness is required. Furthermore, this study showed that farmland could already be identified in April using thermal infrared, multispectral, and LiDAR data. While vegetation indices and texture features for classification relied on multispectral data, SAM segmentation required only RGB data. Therefore, high-time-resolution multispectral data collection in June and July was sufficient for precise planting structure identification. Given that the experimental field selected in this study represents the typical cropping structure of the Loess Plateau [73], the optimal time window identified for fine-scale retrieval of cropping structure and the features extracted from high-resolution data have transferability to crop-growing regions across the Loess Plateau. However, during transfer, attention should be paid to differences in source data resolution and to regional variations in climate, topography, and other factors [74].

The UAV data and synchronous ground survey data also revealed that LiDAR CHM measurements from June to September aligned closely with ground survey data, with

average crop height errors under 5 cm. In contrast, errors exceeded 10 cm in April and May, indicating that CHM accuracy was strongly correlated with crop maturity: the more mature the crop, the more accurate the UAV-derived plant height. A comparison between NDVI values derived from multispectral data and field-measured greenness (Figure 9) showed a high degree of fit across April to September, confirming a positive correlation between vegetation greenness and NDVI.

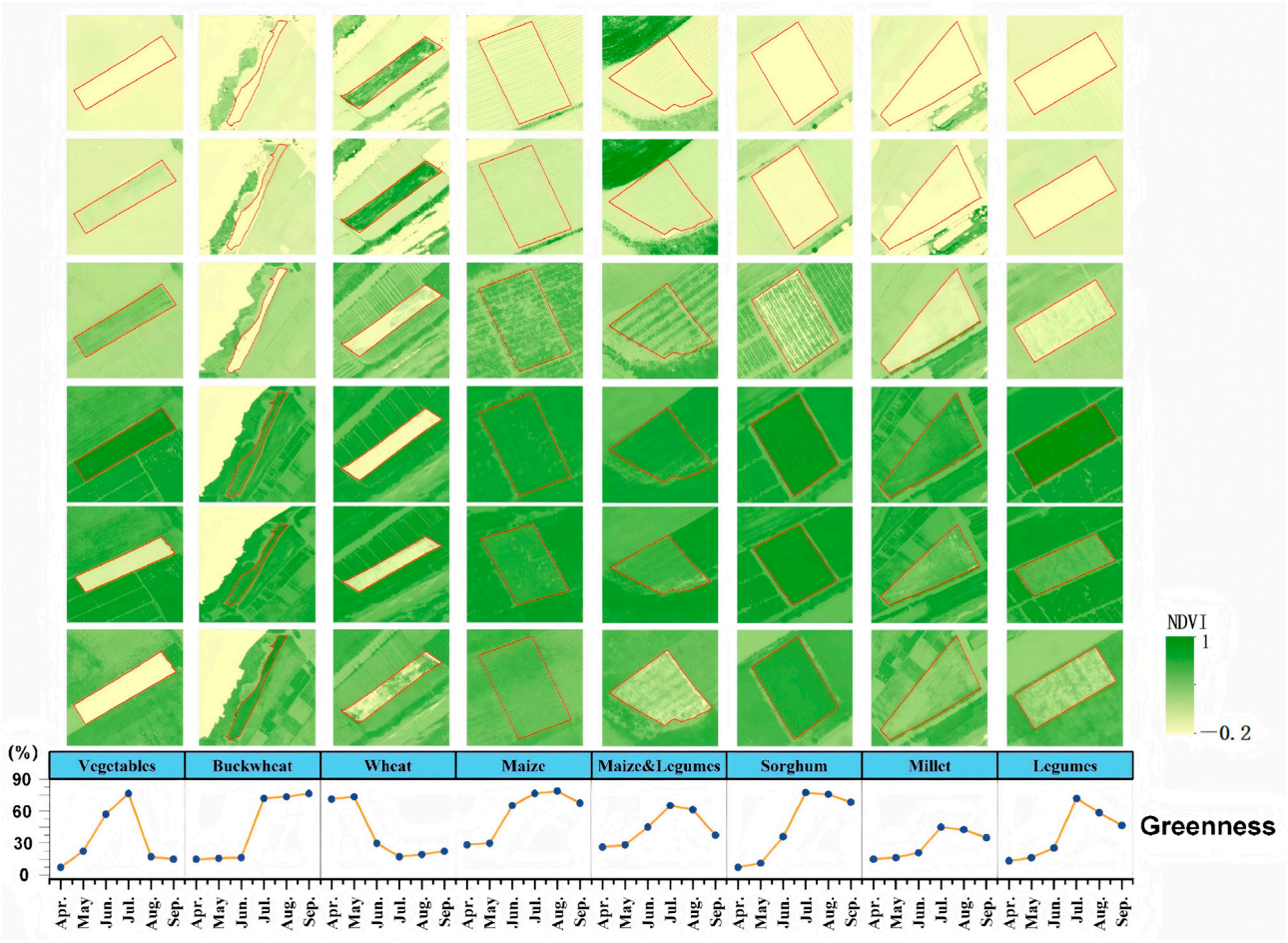


Figure 9. Monthly variation in NDVI and greenness for monoculture structures (April–September).

5.3. Effects of Temporal Phase, Parcel Heterogeneity, and Crop Density on Segmentation

The extraction results were strongly influenced by the segmentation method. As shown in the results, SAM exhibited low recognition accuracy for plots not fully cultivated in April, which limited the effectiveness of plot and planting structure extraction. By May, with more plots cultivated, SAM segmentation improved in accuracy. By June, when plots were almost fully cultivated, SAM delineated plot boundaries precisely, which substantially enhanced planting structure classification accuracy. Note that the above temporal features were derived from multiple acquisitions conducted within a single year over an experimental field in Huachi, Qingyang. Therefore, when generalizing or reusing these findings, regional characteristics and the limitations of single-year data should be taken into account. Beyond temporal features, the timing and conditions of UAV data acquisition must also be considered. Clear weather and mid-day hours are essential for obtaining sharply defined segmentation boundaries, whereas adverse weather markedly degrades radiometric image quality [75]. Meanwhile, beyond environmental and seasonal variability, soil heterogeneity, microclimate, and crop genotype specificity also

affect observational accuracy. Therefore, to enhance the study's generalizability, evidence from long-term (multi-year) studies should also be considered [76].

Compared with other recent attention-based models, SAM adopts a prompt-driven architecture composed of an image encoder (ViT), a prompt encoder, and a lightweight mask decoder. It supports multiple interactive prompts—points, boxes, and masks—enabling zero-shot generalization [24]. By contrast, models such as Mask2Former are built on mask attention, constraining cross-attention within predicted mask regions [77], while SegFormer uses a hierarchical Transformer encoder to extract multi-scale features [78].

In terms of training and generalization, SAM is trained on an ultra-large-scale dataset and thus exhibits strong zero-shot generalization [24], whereas other methods typically rely on task-specific datasets [78]. Regarding computational efficiency, SAM is generally slower at inference because it must process high-resolution image embeddings, while many of the other approaches are comparatively faster [79].

In terms of segmentation performance, comparative experiments revealed that while MRS could delineate ground object boundaries meticulously, it was highly sensitive to internal texture variations within plots, often resulting in over-segmentation of the same plot into multiple objects. In contrast, SAM produced smoother boundaries while better preserving the geometric properties of ground objects. For plots with multiple crop types, SAM clearly defined their boundaries and effectively retained the geometric features of non-vegetation ground objects [80]. From the perspective of intra-parcel heterogeneity, multi-resolution segmentation (MRS) tends to over-segment in highly heterogeneous farmland (e.g., uneven soil texture, crop growth gradients), leading to ambiguous boundaries; in areas with crop density $\leq$ 50%, boundary IoU decreases by 12–18%. By contrast, due to limited sensitivity to local details, SAM exhibits a 21% increase in under-segmentation in low-heterogeneity areas ($\leq$30% coverage) [2]. When within-parcel crop density is high (>80% coverage), SAM leverages global context to capture the continuity of dense canopies and significantly outperforms MRS (Dice 0.92 vs. 0.85) [81]. Under relatively lower density (30–50%), MRS preserves isolated crop patches through multi-scale processing, achieving an IoU of 0.78; SAM, however, tends to overlook small targets, with the omission (miss) rate rising by 35% [82]. The final classification results confirmed that SAM-based segmentation approach demonstrates superior performance in extracting complex, densely planted structures and in recognizing object boundary information [83]. From the standpoint of computational performance and memory footprint, MRS runs efficiently on CPUs, with memory usage scaling approximately linearly with image resolution; on average, a 1024 $\times$ 1024 image requires about 500 MB–1 GB. By contrast, SAM depends on GPU acceleration and is slower than MRS under identical hardware conditions; its ViT-H configuration requires 4–6 GB of VRAM, and a 50% increase in input resolution raises VRAM consumption by roughly 30% [84].

5.4. The Impact of Classification Methods on Classification Results

In this study, CNN, RF, and SVM models were trained based on the optimally selected spectral and texture features to classify crops. RF and SVM are classical methods, well recognized for their ability to determine feature importance and deliver high-precision classification, respectively. CNN, on the other hand, can autonomously learn farmland features and capture complex non-linear relationships. However, CNN requires a large amount of training data, which is difficult to achieve with the limited sample size of UAV data, despite its high spatial resolution. Consequently, CNN produced lower classification accuracy for each combination of period and features (Table 9) compared with RF and SVM. In addition, there are other relatively mature models, such as Transformer-based architectures and XGBoost. However, Transformer-based models typically require larger

training datasets, and training large models from scratch places even higher demands on the data [85]. By contrast, XGBoost is highly sensitive to sample quality, and class imbalance can substantially affect its performance [86]. Using the same SAM segmentation results, machine learning methods achieved classification accuracies exceeding 90% for images from July and September. This confirmed that UAV-based data, coupled with machine learning methods, can provide reliable technical support for precision farming by leveraging UAV's rapid maneuverability [87].

6. Conclusions

(1) In a representative Qingyang (Loess Plateau) site, April enables early-season identification of cropland extent and winter wheat; June allows fine discrimination of monocropped parcels, with April + June together supporting a preliminary wall-to-wall inventory; July yields high-accuracy, full-coverage classification for all crops except buckwheat; and by September, buckwheat likewise attains high-accuracy discrimination.

(2) RF feature selection shows the top-10 cumulative importance peaking in July (72.26%). Pairwise (ratio/difference) features between April and June reach 67.36% (within 5 pp of July), suiting time-critical applications. Crop-specific optimal windows include July for maize, legumes, maize–legume intercropping, sorghum, millet, and vegetables; September for buckwheat; and May for wheat.

(3) Accuracy and mapped areas: April early-season cropland OA = 82.1%; winter wheat 0.33 ha (4.95 mu) at 94%. June SAM-based segmentation reaches 92.8%. In July, full cropping-structure classification achieves OA = 92.66% (Kappa = 0.9143), with mapped areas—millet 7.61 ha (114.17 mu), legumes 7.02 ha (105.36 mu), vegetables 5.33 ha (79.89 mu), sorghum 0.23 ha (3.44 mu), maize 3.59 ha (53.89 mu), maize–legume intercropping 7.01 ha (105.21 mu); buckwheat is 0.69 ha (10.32 mu) at 73% in July and 0.93 ha (13.89 mu) at 98% in September. SAM-based, U-Net-style segmentation mitigates misclassification of fine, fragmented parcels.

(4) Multi-payload UAV data (LiDAR + thermal + multispectral + RGB) enhance cropland extraction; integrating GLCM textures with vegetation indices reduces spectral confusion; and RF performs strongly for UAV-scale mapping and feature prioritization. Given UAVs' rapid, repeatable, low-cost acquisition, deep learning can be further leveraged once data volume suffices for large-scale training.

(5) This study defines temporal windows and feature sets for UAV-based extraction of cropland and cropping structures in a representative Loess Plateau region, achieving high classification accuracy. Limitations remain: an operational pathway for transferring the approach to large-area satellite imagery is not yet specified, and segmentation/classification choices for regional-scale data are unvalidated. Our experiments focus on complex plots with stringent feature weighting; for simpler, conventional parcels, computation should be pared back to avoid redundancy and reduce cost. Future work will port the workflow to satellite sensors, clarify cross-sensor/scale interoperability, and enable early, high-accuracy cropland and cropping-structure mapping across the Loess Plateau at regional scales.

Author Contributions: Conceptualization, Y.Q. and L.W.; methodology, L.W., Y.Q., J.Z. (Juan Zhang), R.Y., J.Z. (Jinlong Zhang) and C.M.; software, L.W., J.Z. (Juan Zhang), R.Y., J.Z. (Jinlong Zhang), C.M. and H.W.; validation, L.W., R.Y. and J.Z. (Juan Zhang); formal analysis, L.W., Y.Q. and R.Y.; investigation, L.W., Y.Q., J.Z. (Juan Zhang), R.Y., J.Z. (Jinlong Zhang) and C.M.; data curation, L.W., R.Y., J.Z. (Juan Zhang) and C.M.; writing—original draft preparation, L.W.; writing—review and editing, Y.Q. and C.M.; visualization, L.W., R.Y., J.Z. (Juan Zhang) and C.M.; supervision, Y.Q.; project

administration, Y.Q.; funding acquisition, Y.Q. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Provincial Industrialization Application Project of China High-Resolution Earth Observation System (CHEOS) of the State Administration of Science, Technology and Industry for National Defense of PRC (Grant No. 92-Y50G34-9001-22/23).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: Data are available upon request due to restrictions (project data privacy). The data presented in this study are available upon request from the corresponding author. The other data and code used in the study have been shared at https://github.com/WangLu199910/PlantingStructure (accessed on 10 September 2025).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

RF	Random Forest
UAVRSS	UAV Remote Sensing System
UAV	Unmanned aerial vehicle
SAM	Segment Anything Model
PDI	Perpendicular Drought Index
RCNN	Recurrent Convolutional Neural Networks
SAR	Synthetic Aperture Radar
CRF	Conditional Random Field
MRS	Multi-resolution Segmentation
ICP	Iterative Closest Point
GLCM	Gray-Level Co-occurrence Matrix
RMSE	Root mean square error
RTK	Real-Time Kinematic
NDVI	Normalized Difference Vegetation Index
EVI	Enhanced Vegetation Index
NDRE	Normalized Difference Red Edge Index
SAVI	Soil-Adjusted Vegetation Index
MSAVI	Modified Soil-Adjusted Vegetation Index
GNDVI	Green Normalized Difference Vegetation Index
RVI	Ratio Vegetation Index
SR	Simple Ratio Index
BNDVI	Blue Normalized Difference Vegetation Index
DVI	Difference Vegetation Index
MNLI	Modified Non-Linear Vegetation Index
VAR	Variance
HOM	Homogeneity
CON	Contrast
DIS	Dissimilarity
ENT	Entropy
ASM	Second Moment
CORR	Correlation
TIR	Thermal Infrared
CNN	Convolutional Neural Network
CHM	Canopy Height Model

OOB	Out-of-bag
OA	Overall Accuracy
UA	User's Accuracy
PA	Producer's Accuracy
FAR	False alarm rate
SVM	Support Vector Machine

References

1. Ajayi, O.G.; Iwendi, E.; Adetunji, O.O. Optimizing crop classification in precision agriculture using AlexNet and high resolution UAV imagery. *Technol. Agron.* **2024**, *4*, e011. [CrossRef]
2. Phang, S.K.; Chiang, T.H.A.; Happonen, A.; Chang, M.M.L. From satellite to UAV-based remote sensing: A review on precision agriculture. *IEEE Access* **2023**, *11*, 127057–127076. [CrossRef]
3. Ji, S.; Zhang, Z.; Zhang, C.; Wei, S.; Lu, M.; Duan, Y. Learning discriminative spatiotemporal features for precise crop classification from multi-temporal satellite images. *Int. J. Remote Sens.* **2020**, *41*, 3162–3174. [CrossRef]
4. Raja, S.Á.; Sawicka, B.; Stamenkovic, Z.; Mariammal, G. Crop prediction based on characteristics of the agricultural environment using various feature selection techniques and classifiers. *IEEE Access* **2022**, *10*, 23625–23641. [CrossRef]
5. Saikhom, V.; Kalita, M. UAV for Remote Sensing Applications: An Analytical Review. In *International Conference on Emerging Global Trends in Engineering and Technology*; Springer: Singapore, 2022; pp. 51–59.
6. De Swaef, T.; Maes, W.H.; Aper, J.; Baert, J.; Cougnon, M.; Reheul, D.; Steppe, K.; Roldán-Ruiz, I.; Lootens, P. Applying RGB-and thermal-based vegetation indices from UAVs for high-throughput field phenotyping of drought tolerance in forage grasses. *Remote Sens.* **2021**, *13*, 147. [CrossRef]
7. Feng, Q.; Yang, J.; Liu, Y.; Ou, C.; Zhu, D.; Niu, B.; Liu, J.; Li, B. Multi-temporal unmanned aerial vehicle remote sensing for vegetable mapping using an attention-based recurrent convolutional neural network. *Remote Sens.* **2020**, *12*, 1668. [CrossRef]
8. Yang, S.; Song, Z.; Yin, H.; Zhang, Z.; Ning, J. Crop classification method of UVA multispectral remote sensing based on deep semantic segmentation. *Trans. Chin. Soc. Agric. Mach.* **2021**, *52*, 185–192.
9. Wang, F.; Yi, Q.; Hu, J.; Xie, L.; Yao, X.; Xu, T.; Zheng, J. Combining spectral and textural information in UAV hyperspectral images to estimate rice grain yield. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *102*, 102397. [CrossRef]
10. Deng, H.; Zhang, W.; Zheng, X.; Zhang, H. Crop classification combining object-oriented method and random forest model using unmanned aerial vehicle (UAV) multispectral image. *Agriculture* **2024**, *14*, 548. [CrossRef]
11. Alvarez-Vanhard, E.; Corpetti, T.; Houet, T. UAV & satellite synergies for optical remote sensing applications: A literature review. *Sci. Remote Sens.* **2021**, *3*, 100019.
12. Chang, B.; Li, F.; Hu, Y.; Yin, H.; Feng, Z.; Zhao, L. Application of UAV remote sensing for vegetation identification: A review and meta-analysis. *Front. Plant Sci.* **2025**, *16*, 1452053. [CrossRef]
13. Ecke, S. Drone Remote Sensing for Forest Health Monitoring. Ph.D. Thesis, Universität Freiburg, Breisgau, Germany, 2025.
14. Javan, F.D.; Samadzadegan, F.; Toosi, A. Air pollution observation—Bridging spaceborne to unmanned airborne remote sensing: A systematic review and meta-analysis. *Air Qual. Atmos. Health* **2025**, *18*, 2481–2549. [CrossRef]
15. Ma, L.; Liu, Y.; Zhang, X.; Ye, Y.; Yin, G.; Johnson, B.A. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS J. Photogramm. Remote Sens.* **2019**, *152*, 166–177. [CrossRef]
16. Kamilaris, A.; Prenafeta-Boldú, F.X. Deep learning in agriculture: A survey. *Comput. Electron. Agric.* **2018**, *147*, 70–90. [CrossRef]
17. Osco, L.P.; Junior, J.M.; Ramos, A.P.M.; de Castro Jorge, L.A.; Fatholahi, S.N.; de Andrade Silva, J.; Matsubara, E.T.; Pistori, H.; Gonçalves, W.N.; Li, J. A review on deep learning in UAV remote sensing. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *102*, 102456. [CrossRef]
18. Rußwurm, M.; Körner, M. Multi-temporal land cover classification with sequential recurrent encoders. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 129. [CrossRef]
19. Wang, X.; Wang, A.; Yi, J.; Song, Y.; Chehri, A. Small object detection based on deep learning for remote sensing: A comprehensive review. *Remote Sens.* **2023**, *15*, 3265. [CrossRef]
20. Garnot, V.S.F.; Landrieu, L.; Giordano, S.; Chehata, N. Satellite image time series classification with pixel-set encoders and temporal self-attention. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 12325–12334.
21. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
22. Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818.

23. Wang, S.; Azzari, G.; Lobell, D.B. Crop type mapping without field-level labels: Random forest transfer and unsupervised clustering techniques. *Remote Sens. Environ.* **2019**, *222*, 303–317. [CrossRef]
24. Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A.C.; Lo, W.-Y. Segment anything. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 2–3 October 2023; pp. 4015–4026.
25. Huang, Z.; Jing, H.; Liu, Y.; Yang, X.; Wang, Z.; Liu, X.; Gao, K.; Luo, H. Segment anything model combined with multi-scale segmentation for extracting complex cultivated land parcels in high-resolution remote sensing images. *Remote Sens.* **2024**, *16*, 3489. [CrossRef]
26. Zhang, E.; Liu, J.; Cao, A.; Sun, Z.; Zhang, H.; Wang, H.; Sun, L.; Song, M. RS-SAM: Integrating multi-scale information for enhanced remote sensing image segmentation. In Proceedings of the Asian Conference on Computer Vision, Hanoi, Vietnam, 8–12 December 2024; pp. 994–1010.
27. Osco, L.P.; Wu, Q.; De Lemos, E.L.; Gonçalves, W.N.; Ramos, A.P.M.; Li, J.; Junior, J.M. The segment anything model (sam) for remote sensing applications: From zero to one shot. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *124*, 103540. [CrossRef]
28. Yang, R.; Qi, Y.; Zhang, H.; Wang, H.; Zhang, J.; Ma, X.; Zhang, J.; Ma, C. A study on the object-based high-resolution remote sensing image classification of crop planting structures in the loess plateau of eastern gansu province. *Remote Sens.* **2024**, *16*, 2479. [CrossRef]
29. Wang, L.; Qi, Y.; Xie, W.; Yang, R.; Wang, X.; Zhou, S.; Dong, Y.; Lian, X. Estimating Gully Erosion Induced by Heavy Rainfall Events Using Stereoscopic Imagery and UAV LiDAR. *Remote Sens.* **2025**, *17*, 3363. [CrossRef]
30. Raza, M.A.; Yasin, H.S.; Gul, H.; Qin, R.; Mohi Ud Din, A.; Khalid, M.H.B.; Hussain, S.; Gitari, H.; Saeed, A.; Wang, J. Maize/soybean strip intercropping produces higher crop yields and saves water under semi-arid conditions. *Front. Plant Sci.* **2022**, *13*, 1006720. [CrossRef]
31. Glenn, E.P.; Huete, A.R.; Nagler, P.L.; Nelson, S.G. Relationship between remotely-sensed vegetation indices, canopy attributes and plant physiological processes: What vegetation indices can and cannot tell us about the landscape. *Sensors* **2008**, *8*, 2136–2160. [CrossRef]
32. Carlson, T.N.; Ripley, D.A. On the relation between NDVI, fractional vegetation cover, and leaf area index. *Remote Sens. Environ.* **1997**, *62*, 241–252. [CrossRef]
33. Peng, X.; Han, W.; Ao, J.; Wang, Y. Assimilation of LAI Derived from UAV Multispectral Data into the SAFY Model to Estimate Maize Yield. *Remote Sens.* **2021**, *13*, 1094. [CrossRef]
34. Fitzgerald, G.; Rodriguez, D.; O'Leary, G. Measuring and predicting canopy nitrogen nutrition in wheat using a spectral index-The canopy chlorophyll content index (CCCI). *Field Crops Res.* **2010**, *116*, 318–324. [CrossRef]
35. Roujean, J.L.; Breon, F.M. Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. *Remote Sens. Environ.* **1995**, *51*, 375–384. [CrossRef]
36. Daughtry, C.S.T.; Walthall, C.L.; Kim, M.S.; de Colstoun, E.B.; McMurtrey, J.E. Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *Remote Sens. Environ.* **2000**, *74*, 229–239. [CrossRef]
37. Gitelson, A.A.; Kaufman, Y.J.; Merzlyak, M.N. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sens. Environ.* **1996**, *58*, 289–298. [CrossRef]
38. Stone, K.H. Aerial photographic interpretation of natural vegetation in the Anchorage area, Alaska. *Geogr. Rev.* **1948**, *38*, 465–474. [CrossRef]
39. Jordan, C.F. Derivation of Leaf-Area Index from Quality of Light on Forest Floor. *Ecology* **1969**, *50*, 663. [CrossRef]
40. Yang, C.; Everitt, J.H.; Bradford, J.M. Airborne hyperspectral imagery and linear spectral unmixing for mapping variation in crop yield. *Precis. Agric.* **2007**, *8*, 279–296. [CrossRef]
41. Gong, P.; Pu, R.L.; Biging, G.S.; Larrieu, M.R. Estimation of forest leaf area index using vegetation indices derived from Hyperion hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 1355–1362. [CrossRef]
42. Haralick, R.M.; Shanmugam, K.; Dinstein, I.H. Textural features for image classification. *IEEE Trans. Syst. Man Cybern.* **2007**, *12*, 610–621. [CrossRef]
43. Li, Z.; Wang, Q.; Xu, H.; Yang, W.; Sun, W. Multi-Source Remote Sensing-Based Reconstruction of Glacier Mass Changes in Southeastern Tibet Since the 21st Century. *EGUsphere* **2025**, *2025*, 1–29.
44. Kettig, R.L. *Computer Classification of Remotely Sensed Multispectral Image Data by Extraction and Classification of Homogeneous Objects*; Purdue University: West Lafayette, IN, USA, 1975.
45. Song, X.; Xie, P.; Sun, W.; Mu, X.; Gao, P. The greening of vegetation on the Loess Plateau has resulted in a northward shift of the vegetation greenness line. *Glob. Planet. Change* **2024**, *237*, 104440. [CrossRef]
46. Antonarakis, A.; Richards, K.S.; Brasington, J.; Bithell, M.; Muller, E. Retrieval of vegetative fluid resistance terms for rigid stems using airborne lidar. *J. Geophys. Res. Biogeosci.* **2008**, *113*, G02S07. [CrossRef]
47. Pereira, L.G.; Fernandez, P.; Mourato, S.; Matos, J.; Mayer, C.; Marques, F. Quality control of outsourced LiDAR data acquired with a UAV: A case study. *Remote Sens.* **2021**, *13*, 419. [CrossRef]

48. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
49. Strobl, C.; Boulesteix, A.-L.; Kneib, T.; Augustin, T.; Zeileis, A. Conditional variable importance for random forests. *BMC Bioinform.* **2008**, *9*, 307. [CrossRef] [PubMed]
50. Awad, M.; Khan, L. Support vector machines. In *Intelligent Information Technologies: Concepts, Methodologies, Tools, and Applications*; IGI Global: Hershey, PA, USA, 2008; pp. 1138–1146.
51. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1097–1105. [CrossRef]
52. Banko, G. *A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data and of Methods Including Remote Sensing Data in Forest Inventory*; International Institute for Applied Systems Analysis: Laxenburg, Austria, 1998.
53. Zhu, X.; Guo, R.; Liu, T.; Xu, K. Crop yield prediction based on agrometeorological indexes and remote sensing data. *Remote Sens.* **2021**, *13*, 2016. [CrossRef]
54. Zhang, D.; Zhang, M.; Lin, F.; Pan, Z.; Jiang, F.; He, L.; Yang, H.; Jin, N. Fast extraction of winter wheat planting area in Huang-Huai-Hai Plain using high-resolution satellite imagery on a cloud computing platform. *Int. J. Agric. Biol. Eng.* **2022**, *15*, 241–250. [CrossRef]
55. Li, Y.; Porto-Neto, L.; McCulloch, R.; McWilliam, S.; Alexandre, P.; Lehnert, S.; Reverter, A.; McDonald, J.; Smith, C. Comparing genomic prediction accuracies for commercial cows' reproductive performance using GA2CAT and two machine learning methods. *Proc. Assoc. Advmt. Anim. Breed. Genet.* **2023**, *25*, 154–157.
56. Sheykhmousa, M.; Mahdianpari, M.; Ghanbari, H.; Mohammadimanesh, F.; Ghamisi, P.; Homayouni, S. Support vector machine versus random forest for remote sensing image classification: A meta-analysis and systematic review. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 6308–6325. [CrossRef]
57. Sawarkar, A.D.; Shrimankar, D.D.; Ali, S.; Agrahari, A.; Singh, L. Bamboo plant classification using deep transfer learning with a majority multiclass voting algorithm. *Appl. Sci.* **2024**, *14*, 1023. [CrossRef]
58. Zhang, C.; Puspitasari, F.D.; Zheng, S.; Li, C.; Qiao, Y.; Kang, T.; Shan, X.; Zhang, C.; Qin, C.; Rameau, F. A survey on segment anything model (sam): Vision foundation model meets prompt engineering. *arXiv* **2023**, arXiv:2306.06211. [CrossRef]
59. Ke, L.; Ye, M.; Danelljan, M.; Tai, Y.-W.; Tang, C.-K.; Yu, F. Segment anything in high quality. *Adv. Neural Inf. Process. Syst.* **2023**, *36*, 29914–29934.
60. Hall, O.; Hay, G.J.; Bouchard, A.; Marceau, D.J. Detecting dominant landscape objects through multiple scales: An integration of object-specific methods and watershed segmentation. *Landsc. Ecol.* **2004**, *19*, 59–76. [CrossRef]
61. Zhang, Y.-L.; Wang, F.-X.; Shock, C.C.; Feng, S.-Y. Modeling the interaction of plastic film mulch and potato canopy growth with soil heat transport in a semiarid area. *Agronomy* **2020**, *10*, 190. [CrossRef]
62. Yang, W.; Li, Z.; Chen, G.; Cui, S.; Wu, Y.; Liu, X.; Meng, W.; Liu, Y.; He, J.; Liu, D. Soybean (*Glycine max* L.) leaf moisture estimation based on multisource unmanned aerial vehicle image feature fusion. *Plants* **2024**, *13*, 1498. [CrossRef]
63. Tan, W.; Yin, Q.; Zhao, H.; Wang, M.; Sun, X.; Cao, H.; Wang, D.; Li, Q. Disruption of chlorophyll metabolism and photosynthetic efficiency in winter jujube (*Ziziphus jujuba*) Induced by *Apolygus lucorum* infestation. *Front. Plant Sci.* **2025**, *16*, 1536534. [CrossRef]
64. Ljubičić, N.; Popović, V.; Kostić, M.; Vukosavljev, M.; Buđen, M.; Stanković, N.; Stevanović, N. The normalized difference red edge index (NDRE) in grain yield and biomass estimation in maize (*Zea mays* L.). In Proceedings of the XV International Scientific Agricultural Symposium Agrosym, Jahorina, Bosnia and Herzegovina, 10–13 October 2024; pp. 373–378.
65. Avtar, R.; Suab, S.A.; Syukur, M.S.; Korom, A.; Umarhadi, D.A.; Yunus, A.P. Assessing the influence of UAV altitude on extracted biophysical parameters of young oil palm. *Remote Sens.* **2020**, *12*, 3030. [CrossRef]
66. Massey, R.; Sankey, T.T.; Congalton, R.G.; Yadav, K.; Thenkabail, P.S.; Ozdogan, M.; Meador, A.J.S. MODIS phenology-derived, multi-year distribution of conterminous US crop types. *Remote Sens. Environ.* **2017**, *198*, 490–503. [CrossRef]
67. Inglada, J.; Vincent, A.; Arias, M.; Marais-Sicre, C. Improved early crop type identification by joint use of high temporal resolution SAR and optical image time series. *Remote Sens.* **2016**, *8*, 362. [CrossRef]
68. Jin, M.; Xu, Q.; Guo, P.; Han, B.; Jin, J. Crop Classification Method from UAV Images based on Object-Oriented Multi-feature Learning. *Remote Sens. Technol. Appl.* **2023**, *38*, 588–598.
69. Ramos, L.T.; Sappa, A.D. Dual-Branch ConvNeXt-Based Network with Attentional Fusion Decoding for Land Cover Classification Using Multispectral Imagery. In Proceedings of the SoutheastCon 2025, Concord, NC, USA, 22–30 March 2025; pp. 187–194.
70. Allu, A.R.; Mesapam, S. Fusion of Satellite and UAV Imagery for Crop Monitoring. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2025**, *X-G-2025*, 71–79. [CrossRef]
71. Zhu, X.; Cai, F.; Tian, J.; Williams, T.K.-A. Spatiotemporal fusion of multisource remote sensing data: Literature survey, taxonomy, principles, applications, and future directions. *Remote Sens.* **2018**, *10*, 527. [CrossRef]
72. Wu, J.; Zheng, D.; Wu, Z.; Song, H.; Zhang, X. Prediction of buckwheat maturity in UAV-RGB images based on recursive feature elimination cross-validation: A case study in Jinzhong, Northern China. *Plants* **2022**, *11*, 3257. [CrossRef] [PubMed]
73. Zhang, Z.; Whish, J.P.; Bell, L.W.; Nan, Z. Forage production, quality and water-use-efficiency of four warm-season annual crops at three sowing times in the Loess Plateau region of China. *Eur. J. Agron.* **2017**, *84*, 84–94.

74. Zhao, X.; Wang, J.; Ding, Y.; Gao, X.; Li, C.; Huang, H.; Gao, X. High-resolution (10 m) dataset of multi-crop planting structure on the Loess Plateau during 2018–2022. *Sci. Data* **2025**, *12*, 1190. [CrossRef]
75. Wierzbicki, D.; Kedzierski, M.; Fryskowska, A. Assesment of the influence of UAV image quality on the orthophoto production. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2015**, *40*, 1–8. [CrossRef]
76. Catania, P.; Ferro, M.V.; Orlando, S.; Vallone, M. Grapevine and cover crop spectral response to evaluate vineyard spatio-temporal variability. *Sci. Hortic.* **2025**, *339*, 113844. [CrossRef]
77. Cheng, B.; Misra, I.; Schwing, A.G.; Kirillov, A.; Girdhar, R. Masked-attention mask transformer for universal image segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 1290–1299.
78. Xie, E.; Wang, W.; Yu, Z.; Anandkumar, A.; Alvarez, J.M.; Luo, P. SegFormer: Simple and efficient design for semantic segmentation with transformers. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 12077–12090.
79. Song, Y.; Pu, B.; Wang, P.; Jiang, H.; Dong, D.; Cao, Y.; Shen, Y. Sam-lightening: A lightweight segment anything model with dilated flash attention to achieve 30×. *arXiv* **2024**, arXiv:2403.09195. [CrossRef]
80. Wu, S.; Su, Y.; Lu, X.; Xu, H.; Kang, S.; Zhang, B.; Hu, Y.; Liu, L. Extraction and Mapping of Cropland Parcels in Typical Regions of Southern China Using Unmanned Aerial Vehicle Multispectral Images and Deep Learning. *Drones* **2023**, *7*, 285. [CrossRef]
81. Li, J.; Feng, Q.; Zhang, J.; Yang, S. EMSAM: Enhanced multi-scale segment anything model for leaf disease segmentation. *Front. Plant Sci.* **2025**, *16*, 1564079. [CrossRef] [PubMed]
82. Ji, W.; Li, J.; Bi, Q.; Liu, T.; Li, W.; Cheng, L. Segment anything is not always perfect: An investigation of sam on different real-world applications. *Mach. Intell. Res.* **2024**, *21*, 617–630. [CrossRef]
83. Xu, W.; Lan, Y.; Li, Y.; Luo, Y.; He, Z. Classification method of cultivated land based on UAV visible light remote sensing. *Int. J. Agric. Biol. Eng.* **2019**, *12*, 103–109. [CrossRef]
84. Gowda, S.N.; Clifton, D.A. Cc-sam: Sam with cross-feature attention and context for ultrasound image segmentation. In Proceedings of the European Conference on Computer Vision, Milan, Italy, 29 September–4 October 2024; pp. 108–124.
85. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.
86. Shwartz-Ziv, R.; Armon, A. Tabular data: Deep learning is not all you need. *Inf. Fusion* **2022**, *81*, 84–90. [CrossRef]
87. Yan, H.; Zhuo, Y.; Li, M.; Wang, Y.; Guo, H.; Wang, J.; Li, C.; Ding, F. Alfalfa yield prediction using machine learning and UAV multispectral remote sensing. *Trans. Chin. Soc. Agric. Eng.* **2022**, *38*, 64–71.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.