remote sensing

MDPI

Article

A Multi-Source Benchmark Dataset for Deep Learning-Based Cropland Parcel Boundary Extraction

Hairong Ma [1,†], Zhiqing Luo [1,†], Jie Hu [2], Bo Guan [1], Mingxue Zheng [1], Baodong Xu [3], Wensen Yu [4] and Pingting Chen [1,*]

[1] Institute of Agricultural Economy and Technology, Hubei Academy of Agricultural Sciences, Wuhan 430064, China; mahairong1008@hbaas.ac.cn (H.M.); luozhq@hbaas.ac.cn (Z.L.); guanbo16@hbaas.ac.cn (B.G.); zhengmingxue1128@hbaas.ac.cn (M.Z.)
[2] School of Artificial Intelligence, Hubei Open University, Wuhan 430074, China; hujie2@hbou.edu.cn
[3] Digital Agriculture Research Institute/College of Resources and Environment, Huazhong Agricultural University, Wuhan 430070, China; xubaodong@mail.hzau.edu.cn
[4] Fujian Key Laboratory of Big Data Application and Intellectualization for Tea Industry, Wuyi University, Nanping 353000, China; npyws@wuyiu.edu.cn
* Correspondence: chenpingting@hbaas.ac.cn
† These authors contributed equally to this work.

Highlights

What are the main findings?

- A multi-source benchmark dataset encompassing over 180,000 expertly refined cropland parcels is constructed.
- Terrain complexity dictates model selection: multi-task semantic segmentation for plains and edge detection for mountains.

What are the implications of the main findings?

- The formulated empirical guidelines can assist researchers and practitioners in selecting optimal remote sensing data.
- The established robust benchmark can accelerate the development of highly generalizable boundary extraction algorithms.

Abstract

The expertly rectified extraction of cropland parcel boundaries is a critical basis for precision agriculture deployment and land resource management. However, the scarcity of large-scale, high-quality, cross-regional reference datasets limits the generalizability of deep learning models. To address this data bottleneck, we created a standardized, multi-source cropland parcel boundary benchmark dataset. The collection, which covers Europe (Denmark, the Netherlands, Spain) and Asia (Central China), encompasses over 180,000 expert-rectified parcels, with average parcel sizes ranging from 0.05 to 10 hectares and spatial resolutions spanning 0.2 to 10 m. Based on this dataset, we evaluated nine representative deep learning models belonging to three architectural paradigms (edge detection, multi-task semantic segmentation, and instance segmentation) and analyzed the effects of spatial resolution, parcel scale, and topographic relief on extraction accuracy. The experimental results indicate the following: (1) Spatial resolution exerts a non-linear influence on boundary extraction fidelity, where some medium-resolution data conditions (such as 4.8 m imagery in large-scale fields) retained competitive boundary extraction performance for certain models. (2) Even small but regular parcels of merely 0.22 hectares can be extracted with high accuracy, achieving a maximum F1 score of 86.58%. (3) Topographic variations lead to contrasting model performance. In plain and hilly terrain, multi-task semantic segmentation

Check for updates

Academic Editor: Zhenwei Shi

Received: 6 September 2026
Revised: 23 October 2026
Accepted: 24 October 2026
Published: 28 October 2026

Copyright: © 2026 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.

models exhibit obvious advantages. BsiNet yields the highest F1 score of 86.58%, but this value drops to 36.56% in complex mountainous regions. By comparison, edge detection models perform better in mountainous areas, where RCF attains the maximum F1 of 70.85%. In summary, this study establishes a standardized, expertly refined reference parcel dataset and provides an empirical evaluation of how different model architectures perform across varied observational settings, offering an empirical baseline for future model development and evaluation in cropland boundary extraction.

Keywords: cropland parcel boundary extraction; benchmark dataset; multi-source remote sensing; deep learning

1. Introduction

The timely and accurate acquisition of agricultural data is fundamental to advancing agricultural modernization. Among agricultural data, cropland information is particularly pivotal, representing one of the most essential data types within the agricultural sector. Cropland parcels, as the smallest operational units of agricultural production and farm management, require the precise delineation of their boundaries, which remains a cornerstone of precision agriculture. Accurate boundary information enables agricultural authorities to manage land resources more efficiently, monitor crop conditions, assess the impacts of disasters, and develop effective policies to safeguard food security [1,2]. Specifically, detailed cropland parcel boundary data support the development of foundational cropland statistics [3], the creation of cadastral maps [4], the construction of regional agricultural models [5], and the establishment of comprehensive cropland databases [6]. Additionally, these data play a critical role in applications such as crop yield prediction [7], crop classification and identification [8], crop distribution mapping [9], pest and disease management [10], precision fertilization [11], and the autonomous navigation of agricultural machinery [12]. Furthermore, they contribute substantially to regional agricultural management [13], land use planning [14], and water resource management [15]. Thus, the accurate and efficient extraction of cropland parcel boundaries over large areas holds significant scientific and practical value for improving agricultural productivity and supporting long-term sustainability.

Remote sensing technology has emerged as a powerful tool for acquiring agricultural information, including estimates of cropland areas [16,17], crop distribution patterns [18,19], and crop growth status [20,21]. The methods for extracting cropland information from remote sensing data have undergone significant evolution, progressing from early human–computer interactive visual techniques [22], pixel-based texture analysis [23], and object-oriented approaches [24], to more advanced machine learning frameworks [25–27].

Traditional visual interpretation, which relies on expert-driven manual annotation of high-resolution imagery informed by prior knowledge, has demonstrated relatively high accuracy for delineating cropland parcel boundaries. However, as the scale and volume of cropland data continue to expand, these manual methods become increasingly labor-intensive and time- and resource-consuming. Consequently, automated approaches for cropland parcel boundary extraction have grown in importance. Deep learning, with its strong capacity to learn complex spatial patterns and autonomously extract hierarchical features, has achieved substantial progress in computer vision such as cropland parcel boundary extraction. The superior performance and broad potential of deep learning techniques in remote sensing image processing have accelerated their application in cropland parcel boundary extraction, resulting in enhanced robustness and improved accuracy [28,29]. As

a result, deep learning-based methods for cropland parcel boundary extraction have become a prominent and rapidly growing area of research in the remote sensing community.

In the context of cropland parcel boundary extraction, current deep learning models can be broadly categorized into four main types: edge detection, multi-task learning, semantic segmentation, and instance segmentation networks. Edge detection models such as RCF [30,31], HED [32,33], and D-LinkNet [34] enhance fine-grained edge feature extraction by leveraging multiscale semantic cues. Although these networks achieve strong performance in general edge detection tasks, they are not specifically optimized for identifying cropland parcel boundaries. As a result, their outputs often contain irrelevant or noisy edges that degrade the precision and geometric accuracy of the extracted parcel boundaries. To address the aforementioned shortcomings, multi-task learning networks combine the strengths of segmentation and boundary detection. Models such as cascaded architectures [35,36], BSNet [37], BsiNet [38], ResUNet-a [39,40], SEANet [41], and TFNet [42] jointly optimize region-level and edge-level predictions, thereby improving boundary continuity and semantic coherence. For example, the Dual Spatio-Temporal Fusion Network (DSTFNet) integrates high-resolution GF-2 imagery with multi-temporal Sentinel-2 data to exploit both fine-grained spatial information and spectral–temporal dynamics [43]. Nevertheless, most multi-task learning approaches still depend on postprocessing operations (e.g., watershed or morphological segmentation) to obtain parcel instances, which may introduce boundary distortions and lead to under-segmentation of small parcels. Semantic segmentation models such as DeepLab [44], DenseNet [45], CNN-based models [46], UNet [47], SegNet [48], and Fully Convolutional Networks (FCNs) [49,50] are widely used for delineating cropland from remote sensing imagery. These models perform dense per-pixel classification but often experience spatial information loss due to repeated downsampling operations, which is particularly critical for accurate boundary delineation. This limitation can result in discontinuous or imprecise parcel boundaries, and ambiguities between adjacent parcels may lead to parcel adhesion and incomplete vectorization. Instance segmentation models including Mask R-CNN [51], SOLOv1 [52], DeepSnake [53], and E2EC [54] directly generate discrete parcel instances by modeling object contours. Such approaches maintain topological independence between parcels and produce smoother, more accurate boundaries. For example, the E2EVAP model [55], which integrates DeepSnake and E2EC, demonstrates superior capability in preserving fine-scale geometry for extracting fragmented parcels from high-resolution satellite imagery.

As a data-driven approach, the performance of deep learning models heavily relies on the quantity and quality of annotated training data. However, publicly available datasets for cropland information extraction remain scarce, and many studies require manual re-annotation of raw data. This process not only increases research time and costs but also hinders the widespread adoption and scalability of deep learning techniques in agricultural applications. The publicly available cropland parcel datasets can be classified into several categories, including (i) datasets based on Jilin-1 optical remote sensing imagery for specific regions of China [56]; (ii) datasets derived from Sentinel-2 satellite and aerial imagery for European regions [57]; (iii) datasets covering parts of France and India, sourced from Systeme Probatoire d' Observation de la Terre and PlanetScope optical remote sensing imagery [58]; (iv) the Agriculture-Vision dataset, which includes aerial optical remote sensing images from U.S. croplands and uses red, green, blue (RGB) and near-infrared (NIR) channels for semantic segmentation tasks [59]; and (v) the Panoptic Agricultural Satellite Time Series (PASTIS) dataset, which integrates time-series Sentinel-2 multispectral imagery and Sentinel-1 radar imagery for French cropland parcels, supporting both semantic and instance segmentation tasks [60].

Despite their utility, these public datasets exhibit several notable limitations that impede both research progress and practical implementation. First, inadequate spatial resolution: Some datasets rely on remote sensing imagery with insufficient spatial resolution, limiting their ability to capture the fine-scale detail needed for accurate cropland parcel boundary extraction. Second, incomplete annotations: Some datasets contain incomplete or inaccurate parcel boundary annotations, which negatively affect the training process and reduce the overall performance of deep learning models. Third, limited parcel diversity: Many datasets are geographically or thematically constrained, often focusing on specific regions or crop systems. This limited diversity restricts model generalizability and weakens transfer learning performance. Finally, limited remote sensing sources: Most datasets rely on a narrow range of remote sensing imagery, often from only one or two sensor types, limiting the potential to leverage multi-source data for training more robust models.

To address the aforementioned challenges, this study makes the following key contributions:

1. A benchmark dataset containing over 180,000 unique physical cropland parcels was constructed. The dataset is derived from multi-source remote sensing imagery covering diverse topographic landscapes, and expert manual rectification was applied to minimize spatiotemporal misalignments and label noise.

2. The extraction performance of nine representative deep learning models was analyzed across empirical settings of spatial resolution, parcel scale, and terrain undulation, and potential physical and geometric influences were identified. These findings provide preliminary empirical insights to support future algorithm development and evaluation in cropland parcel delineation.

2. Materials and Methods

2.1. Sample Area

Six sampling areas across Europe and China were selected to evaluate deep learning model performance across diverse gradients of cropland parcel morphology, topography, and spatial resolution. Specifically, the European areas are situated in predominantly flat lowland landscapes, facilitating the comparative observation of parcel size and geometric regularity with minimal topographic variation. In parallel, the Chinese areas encompass plains, hills, and mountains within Hubei Province, establishing an empirical topographic gradient to analyze model sensitivity across varying terrain complexities under a broadly consistent regional climate.

2.1.1. European Sampling Areas

To represent Europe's highly land-intensive and mechanized agricultural landscape, this study selected Denmark, the Netherlands, and Spain as the study areas. The cropland parcels in these areas are generally contiguous and well defined, with regular shapes, and the average area of a single parcel ranges from 1.5 to 10 hm^2. Figure 1 shows the geographic locations of the European study areas and the spatial distribution of the random sampling grid.

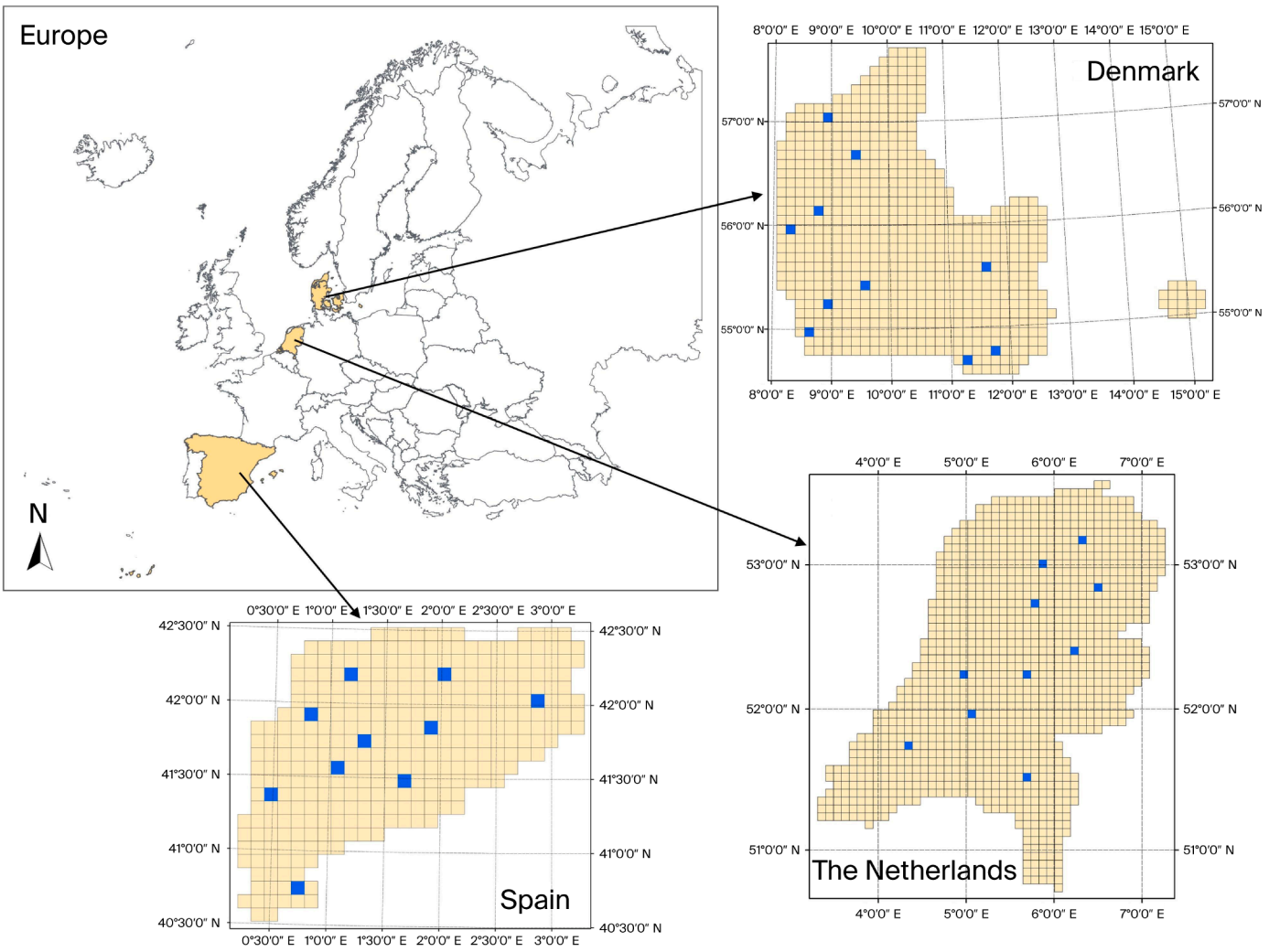


Figure 1. The geographic locations of the European study areas and the spatial distribution of the random sampling grid. Note: The map is intended solely to illustrate the spatial distribution of the sampling grids and does not represent official national borders.

(1) Denmark

Denmark is a prime example of intensive agriculture in Northern Europe. Approximately 60% of its land area (a total of about 42,900 km^2) is designated as cropland, and generally exhibits a spatial pattern characterized by contiguous, concentrated blocks with relatively large individual parcel scales. At the morphological level, while the boundaries of some local parcels are straight and their contours are distinct, the cropland parcels across the region as a whole are predominantly irregular composite polygons (with trapezoids and various irregular polygons intermingled), exhibiting a high degree of randomness in their spatial orientation. The average parcel area in this region is approximately 6 hm^2. This unique spatial pattern reflects a dynamic balance between modern, highly mechanized agriculture and the legacy features of historical land use. This dual characteristic, which clearly defines local boundaries and a complex overall geometric configuration, can be used to test the geometric perception capabilities and multiscale generalization abilities of deep learning models.

(2) Netherlands

The Netherlands, located in Western Europe, has a total land area of approximately 41,500 km^2, with agricultural and pastoral land accounting for about 55% of the total area.

This region follows a typical European model of intensive farming, characterized by highly concentrated clusters of croplands. The parcels have regular geometric shapes, generally consisting of narrow rectangles with a high length-to-width ratio, and are uniformly oriented. The average parcel scale is approximately 10 hm^2. This cropland landscape is the result of the combined effects of large-scale land leveling, highly mechanized farming, and precision agriculture. For deep learning models, such narrow and highly regular parcels effectively test the network's ability to maintain boundary continuity over long distances and preserve geometric details such as right-angle corners and long, straight boundaries.

(3) Spain

Spain, located in southwestern Europe, has a total land area of approximately 506,000 km^2, of which 37% is cropland. Agricultural production is primarily distributed across gently sloping, low-altitude plains and is concentrated in provinces such as Barcelona, Girona, Lleida, and Tarragona. Unlike the highly contiguous spatial layout of cropland in Northern and Western Europe, Spain's cropland exhibits significant spatial heterogeneity. Parcel continuity within the region is relatively weak, as parcels are frequently fragmented by natural vegetation, topographic undulations, roads, and waterways, ultimately forming a patchy, semi-fragmented pattern where contiguous parcels and scattered, fragmented parcels are intermixed. In terms of morphology, cropland parcels in this region are characterized by pronounced geometric irregularity, primarily consisting of various irregular polygons. The average parcel area in the region is approximately 1.5 hm^2, reflecting a moderately refined farming model adapted to complex topography. For deep learning models, this highly heterogeneous scenario presents a significant testing challenge, allowing for the evaluation of a network's ability to identify ambiguous parcel boundaries amid complex background interference—such as vegetation—and topological noise.

2.1.2. Chinese Sampling Areas

Hubei Province in Central China was selected because it encompasses plains, hills, and mountainous terrain within a broadly consistent subtropical monsoon climate. This regional setting facilitates the comparative observation of topographic variations while keeping the macro-climatic and cropping background relatively uniform. Furthermore, while large-scale regular agriculture is represented by the European sampling areas in our dataset, the areas in Hubei capture typical Asian smallholder agriculture characterized by high parcel fragmentation, dense boundaries, and micro-scale cropland configurations (0.05–0.22 hm^2), providing an empirical contrast in parcel scale and morphology.

(1) Jiangxia District

This sampling area spans roughly 238 km^2 and consists mostly of gentle plains. The average area of cropland parcels in the region is only 0.22 hm^2. Geometrically, the parcel outlines are rather regular, consisting primarily of small rectangles and narrow, elongated strips. This fine-grained division of parcels produces an extraordinarily dense surface spatial texture, resulting in an ultra-dense, fine-grained agricultural landscape. This complicated scenario effectively evaluates deep learning models' capacity to delineate dense parallel boundaries while suppressing flaws such as the merging of narrow, nearby parcels and under-segmentation.

(2) Echeng District

This sampling area is roughly 78 km^2 and has a typical hilly environment. Due to the undulating landscape, cropland is primarily scattered along moderate slopes and valleys. Natural vegetation and water networks create a dispersed, fragmented pattern, with an average parcel scale of only 0.05 hm^2. In terms of geometric morphology, to adhere to local micro-topography and terrain contours, the parcels are highly irregular, consisting

primarily of terraced croplands and complicated polygons with meandering boundaries. This agricultural landscape provides an excellent environment for thoroughly testing deep learning models. It particularly evaluates the models' ability to extract curved and non-orthogonal contours, as well as their stability in detecting low-contrast boundaries (such as terraced cropland ridges) in the presence of terrain-induced shading and occlusion.

(3) Tongshan County

This sampling area is mountainous, covering approximately 2680 km^2. Steep slopes and sharply carved valleys significantly reduce cropland parcels' spatial continuity. They are found on gentle valley floors and hillsides, and are mostly made up of tiny terraces and polygons that follow contour lines with meandering and uneven boundaries. The average parcel scale is just 0.05 hm^2. This geographical pattern clearly displays the high-altitude environment's substantial spatial limits on cropland distribution. The high degree of parcel isolation and the undulating topography in this scene provide an effective test bed for deep learning models to assess their ability to identify sparse, minute cropland targets amid strong background noise such as forests and under conditions of dramatic fluctuations in illumination caused by mountain shadows.

2.2. Dataset Sources

To ensure the diversity of the remote sensing datasets and support multi-perspective, multi-dimensional analyses for cropland parcel boundary extraction, this study adopted multi-source and multi-resolution remote sensing imagery covering six study areas. The benchmark encompasses 187,256 unique physical parcels and more than 280,000 multi-source parcel observations and was divided into nine data subsets, each containing approximately 10,000 parcels.

Uniform random grid sampling was utilized in Europe to ensure a balanced spatial distribution of the sample. Ten grids, each measuring 10 km × 10 km (Figure 1), were established in Denmark, the Netherlands, and Spain. Our sampling grids captured 12,015, 60,177, and 45,463 unique physical cropland parcels in Denmark, the Netherlands, and Spain, respectively (detailed in Table 1). For Denmark and Spain, high-resolution remote sensing imagery with a resolution of 0.5 m from Global Mapper was chosen; for the Netherlands, we used multiple imagery sources including Global Mapper (0.6 m), Geospatial Earth (2.4 m), Pleiades (4.8 m), and Sentinel-2 (10 m), yielding over 170,000 multi-sensor parcel observations.

Twelve grids were selected from the Chinese sampling areas across Jiangxia District (plains), Echeng District (hills), and Tongshan County (mountains) (Figure 2). Rather than an unconstrained random spatial sampling scheme, the locations of the Unmanned Aerial Vehicle (UAV) survey grids were determined based on targeted aerial photogrammetric campaigns from concurrent agricultural survey projects, reflecting practical logistical constraints and regional farming priorities. These sampling units comprise four 5 km × 5 km grids in Tongshan alongside four 2 km × 2 km grids each in Jiangxia and Echeng. Following data screening and quality verification, Jiangxia, Echeng, and Tongshan areas yielded 10,443, 15,779, and 35,376 unique physical parcels, respectively. Regarding data sources, 0.5 m Global Mapper orthomosaics were employed in the plain areas; 0.2 m resolution UAV orthomosaics were utilized across the hilly sites; and in the mountainous areas, 0.5 m UAV orthomosaics were used as the primary benchmark data.

Table 1. Comprehensive specifications and metadata for all remote sensing imagery across the six study regions.

Study Areas	Geomorphology	Maximum Elevation/Minimum Elevation/Avg Elevation (m)	Platform/Source	Sampling Scheme	Acquisition Dates	Spatial Resolution (m)	Total Sampling Area (km^2)	Avg Parcel Scale (hm^2)	Sampled Unique Parcels
Denmark (Northern Europe)	Plains	171/−7/33	Global Mapper	10 grids (10 km × 10 km)	October 2022	0.5	1000	6	12,015
The Netherlands (Western Europe)	Plains	322/−7/30	Global Mapper	10 grids (10 km × 10 km)	October 2022	0.6	1000	10	8003
			Geospatial Earth	A rectangular area located in the central Netherlands	Summer 2022	2.4	5331	10	32,599
			Pleiades	Three regions located in western, central, and southern parts of the Netherlands	Summer 2022	4.8	11,060	10	63,665
			Sentinel-2	Three regions located in western, central, and southern parts of the Netherlands	Summer 2022	10	11,060	10	63,665
Spain (Southwestern Europe)	Plains and Hills	3143/0/600	Global Mapper	10 grids (10 km × 10 km)	October 2022	0.5	Plains (500) Hills (500)	1.5	Plains (23,902) Hills (21,561)
Jiangxia District (Hubei Province, China)	Plains	67/11/24.5	Global Mapper	4 grids (2 km × 2 km)	October 2022	0.5	16	0.22	10,443
Echeng District (Hubei Province, China)	Rolling Hills	485/15/36	UAV Platforms	4 grids (2 km × 2 km)	2023	0.2	16	0.05	15,779
Tongshan County (Hubei Province, China)	Steep Mountains	1657/20/800	UAV Platforms	4 grids (5 km × 5 km)	July 2023/December 2023/May 2024	0.5	100	0.05	35,376

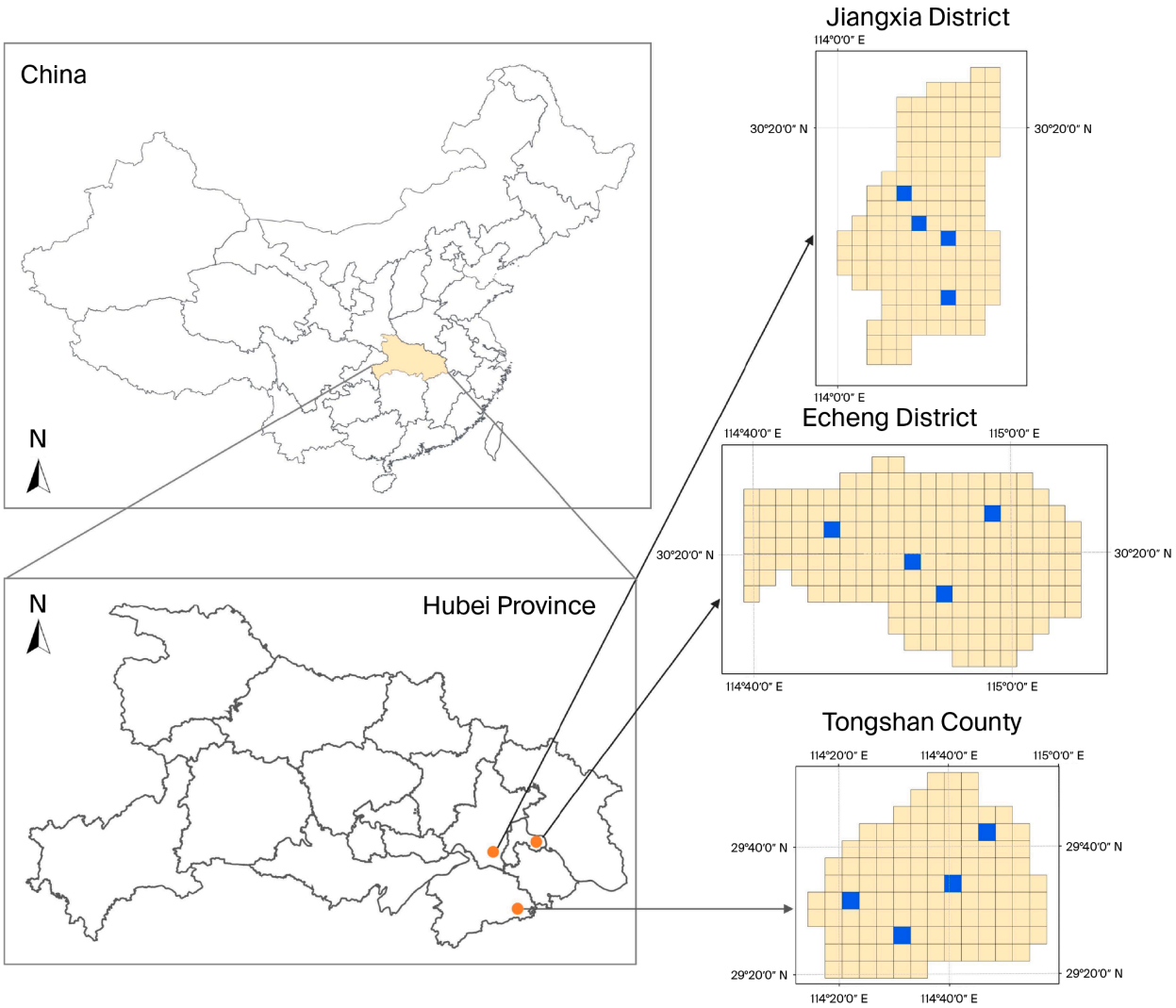


Figure 2. The geographic locations of the Chinese study areas and the spatial distribution of the selected UAV survey grids.

All benchmark imagery adheres to uniform technical specifications: it comprises three spectral channels (RGB) with a radiometric bit depth of 8 bits; all image samples are screened to ensure cloud-free and fog-free imaging conditions; no additional secondary resampling, pan-sharpening, or lossy re-compression was introduced by the authors beyond the standard product generation pipelines delivered by the data provider. The three international study areas (the Netherlands, Denmark, and Spain) adopt the WGS 1984 coordinate system, while the Chinese study areas employ the CGCS 2000 coordinate system. The data sources and preprocessing workflows vary across regions and resolution tiers: the 0.5 m and 0.6 m high-resolution imagery for the Netherlands, Denmark, and Spain sites are standard geometrically compliant products sourced from the Global Mapper platform, with no additional secondary processing applied to the as-delivered data; the medium- and low-resolution imagery (2.4 m, 4.8 m, and 10 m) for the Netherlands site underwent standard preprocessing workflows including radiometric correction, orthorectification, and atmospheric correction; for the Chinese study areas, imagery over Jiangxia was acquired by the Global Mapper satellite constellation and underwent the same standard preprocessing of radiometric correction, orthorectification, and atmospheric correction, while imagery for the Echeng and Tongshan was collected by UAV photogrammetric systems, and the final orthoimage products were processed via color balancing and mosaicking. For quantitative

terrain characterization, regional elevation attributes were extracted from the Shuttle Radar Topography Mission 1 Arc-Second Global DEM (nominal 30 m spatial resolution), jointly acquired by NASA and the U.S. National Geospatial-Intelligence Agency under the Department of Defense, with an absolute vertical accuracy of LE90 < 16.0 m and an absolute horizontal accuracy of CE90 < 20.0 m. Exhaustive technical specifications are systematically documented in Table 1.

2.3. Parcel Boundary Rectification

To rectify these geometric errors and prevent label noise from propagating into model training, this study implemented an expertly refined geometric refinement protocol executed by two Ph.D. researchers specializing in agricultural remote sensing and Geographic Information System, each possessing over three years of experience in satellite and UAV image interpretation. The sampling areas were partitioned between the researchers for primary manual rectification. Operators adjusted vector vertices to align with visible physical crop boundaries, digitally reconstructing omitted boundaries (e.g., newly partitioned plots) and manually excluding non-agricultural features (e.g., farm roads, ditches, or rural buildings). Standard GIS topological rules were enforced during editing to eliminate polygon overlaps, unclosed dangles, and sliver gaps.

To ensure label consistency and verify annotation repeatability across all sampling areas, a stratified random sample of 5–10% of the parcels underwent double-blind cross-validation. Boundary discrepancies between the two annotators exceeding an operational spatial threshold of 2 pixels (IoU (Intersection over Union) < 0.90) were flagged and resolved through joint consensus meetings by cross-referencing multi-temporal high-resolution imagery and local contextual features.

In publicly accessible international agricultural databases (such as European Land Parcel Identification System platforms), parcel vectors are frequently based on self-reported declarations submitted by farmers for subsidy administration rather than exhaustive geodetic surveys. Consequently, raw cadastral layers often exhibit sporadic label omissions, coarse boundary approximations, or administrative delineations that do not align with actual physical field margins. To address both farmer-declared reporting discrepancies and temporal drift, the expertly refined workflow utilized raw cadastral vectors solely as initial topological priors. Remote sensing interpretation experts systematically audited each polygon against contemporaneous high-resolution orthomosaics, correcting erroneous or missing boundaries and updating any parcels that underwent structural land use changes. Figure 3 shows a comparison of vector details before and after rectification.

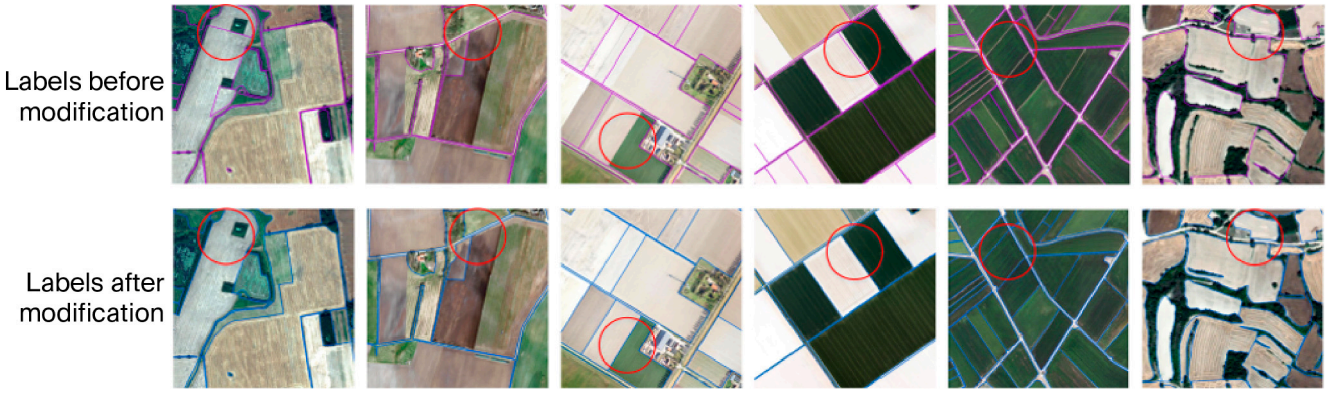


Figure 3. Comparative analysis of cropland parcel labels before and after data standardization.

2.4. Image Chip Generation

To ensure methodological clarity and transparent data accounting, we distinguish between three core structural dataset entities: (1) unique physical parcels, representing distinct geographic entities on the ground delineated with unique cadastral boundary IDs; (2) multi-sensor parcel observations, representing individual instances where a physical parcel is observed across different sensor platforms, spatial resolution tiers; and (3) image chips, representing the standardized, non-overlapping raster tiles generated to serve as input tensors for the 7:1:2 train/validation/test split. Given the substantial computational overhead of training and benchmarking multiple deep learning models, processing all collected imagery would exceed feasible workstation computing capacities. Therefore, 1–2 representative large-scale remote sensing scenes were selected from each study area (totaling eleven primary image scenes across the six regions, as detailed in Table 2). These selected large-scale scenes were subsequently cropped into non-overlapping image chips with dimensions tailored to the specific input requirements of each evaluated deep learning model, with detailed patch configurations and chip statistics provided in Table 2 below.

Table 2. Quantitative summary of image dimensions, spatial resolution and numbers of image chips for model training and evaluation across study areas.

Sampling Area	Spatial Resolution	Image Dimensions (Pixels)	HED 512 × 512 (Image Chips)	RCF 512 × 512 (Image Chips)	D-LinkNet 1024 × 1024 (Image Chips)	BsiNet 512 × 512 (Image Chips)	ResUNet-a 224 × 224 (Image Chips)	TFNet 512 × 512 (Image Chips)	E2EC 640 × 640 (Image Chips)	SOLOv2 640 × 640 (Image Chips)	YOLOv8-Seg 640 × 640 (Image Chips)
Denmark	0.5 m	29,495 × 28,885	3192	3192	784	3192	16,768	3192	2070	2070	2070
Netherlands	0.6 m	16,667 × 16,668	1024	1024	256	1024	5476	1024	676	676	676
Netherlands	2.4 m	22,940 × 24,920	2112	2112	528	2112	11,322	2112	1330	1330	1330
Netherlands	4.8 m	11,508 × 12,474	528	528	132	528	2805	528	323	323	323
Netherlands	10 m	5452 × 5959	110	110	25	110	624	110	72	72	72
Spain	0.5 m	27,356 × 27,458	2809	2809	676	2809	14,884	2809	1764	1764	1764
Jiangxia	0.5 m	9210 × 8762 9204 × 8757	578	578	128	578	3198	578	364	364	364
Echeng	0.2 m	10,001 × 10,000 10,001 × 10,000	722	722	162	722	3872	722	450	450	450
Tongshan	0.5 m	10,002 × 10,001	361	361	81	361	1936	361	225	225	225

As demonstrated in Figure 4, the parcel samples collected from six sampling areas shared common traits while also differing significantly.

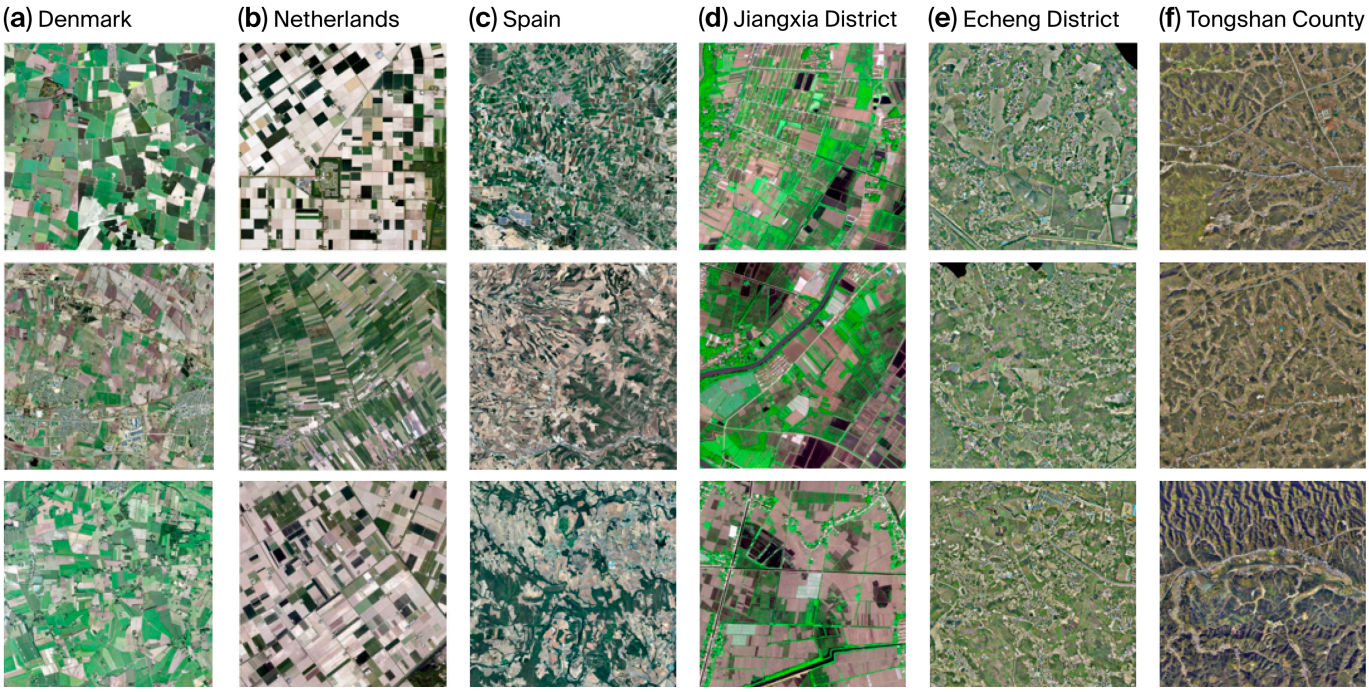


Figure 4. Representative original images of cropland parcels from six sampling areas.

2.5. Deep Learning Models

To explore how deep learning models respond to different types of remote sensing imagery and landscape conditions, this study evaluated nine representative models spanning three mainstream architectural paradigms: edge detection networks, multi-task semantic segmentation networks, and instance segmentation networks.

Edge detection networks: HED and RCF are built upon Fully Convolutional Network topologies. HED adopts deep supervision, while RCF leverages hierarchical feature aggregation to mitigate boundary blurring in target contour extraction. This study also includes D-LinkNet, a model originally developed for road extraction. This encoder–decoder architecture employs dilated convolutions to enlarge the receptive field and shows favorable adaptability for small, continuous boundaries of cropland parcels.

Multi-task semantic segmentation networks alleviate geometric artifacts common to conventional pixel-wise classification, such as object merging, missed small objects, and topological distortion, through joint optimization of region segmentation, boundary extraction and distance field prediction. We adopt ResUNet-a, a well-established multi-task segmentation architecture for high-resolution remote sensing imagery. Using a UNet backbone, it combines dilated convolutions and Pyramid Scene Parsing (PSP) Pooling for sequential multi-task inference of parcel masks and boundaries. BsiNet, a specialized multi-task network for cropland boundary extraction from high-resolution satellite images, is also evaluated. It uses three parallel decoders within a unified architecture to predict segmentation masks, border maps, and distance fields. We also evaluated TFNet. This model improves multi-task parcel delineation by decoupling feature representation into mask and boundary branches, and integrates boundary priors from the Segment Anything Model (SAM) to better capture faint boundaries across heterogeneous agricultural landscapes.

Instance segmentation networks: Unlike semantic segmentation, instance segmentation models assign unique identities to individual objects, thereby mitigating parcel adhesion. We adopt E2EC, built on the DeepSnake architecture. It leverages contour initialization and dynamic matching loss constraints to maintain geometric integrity for object contours. We also evaluate SOLOv2 [61], which employs Matrix Non-Maximum Suppression (Matrix

NMS) and delivers high inference efficiency for densely packed parcels. In addition, we include the segmentation variant of YOLOv8 [62] to assess the trade-off between real-time deployment efficiency and expert-rectified contour delineation of this cutting-edge model.

2.6. Data Splitting and Training Configuration

All models were implemented on a workstation equipped with a single NVIDIA Quadro P4000 GPU (12GB of VRAM) (NVIDIA, Santa Clara, CA, USA), Intel Xeon Bronze 3104 CPU (1.70 GHz) (Intel, Santa Clara, CA, USA), and 32 GB RAM, utilizing PyTorch 1.13.1 and CUDA 11.6. A fixed global random seed (seed = 42) and CuDNN computation backend flags were applied across all training pipelines.

Partitioning of the remote sensing imagery was conducted at the level of independent, non-overlapping image chips rather than individual vector parcels, thereby preventing spatial data leakage across splits. To accommodate the varying architectural complexities, receptive cropland sizes, and input requirements of different deep learning models without causing aspect ratio distortion, the cropping dimensions were tailored specifically for each model. Subsequently, these non-overlapping image chips were randomly assigned to the training, validation, and test sets in a 7:1:2 ratio.

To align with the original implementations and established practices in the literature, model-specific pretrained weights and initialization strategies were retained where applicable. D-LinkNet employed a ResNet-34 backbone initialized with ImageNet-1K V1 pretrained weights, whereas E2EC, SOLOv2, and YOLOv8-seg were initialized using COCO-pretrained weights. HED and RCF were initialized from the pretrained checkpoints provided by their respective official implementations. BsiNet was initialized from the pretrained checkpoint supplied with its implementation and subsequently fine-tuned on the proposed dataset. TFNet used the multiscale features of pretrained FastSAM-x as the visual encoder. During training, the encoder parameters are frozen, while the convolutional layers and linear layers in the remaining adaptation, feature interaction and decoding modules are initialized using Kaiming normal initialization. In contrast, ResUNet-a was trained from scratch without pretrained weights, with convolutional layers initialized using the normal initialization strategy adopted in the original implementation.

For rigorous and unbiased evaluation, all benchmarked architectures retained their native network configurations and employed the default loss formulations specified in their foundational publications. HED used weighted binary cross-entropy (BCE) for multiple side outputs; RCF used class-balanced BCE across its edge prediction outputs; D-LinkNet used BCE + Soft Dice loss; ResUNet-a used Tversky loss; BsiNet used cross-entropy (CE)/negative log-likelihood (NLL) losses for foreground and contour prediction together with mean squared error (MSE) for the distance map; E2EC used a modified Focal loss for the center heatmap and Smooth L1 losses for contour regression and terminal contour matching; Focal loss + Dice loss for SOLOv2; and Task-Aligned Focal loss + Complete IoU (CIoU) + Distribution Focal Loss (DFL) + BCE for YOLOv8-seg. TFNet employs binary cross-entropy loss for the region branch, class-balanced weighted binary cross-entropy loss for the boundary branch, and a cosine similarity-based region–boundary consistency loss. The total loss is defined as the sum of these three loss terms. The comprehensive hyperparameter matrix is summarized in Table 3.

Table 3. Detailed experimental protocols, loss function formulations, and hyperparameter matrices ensuring benchmark reproducibility.

Model	Paradigm	Backbone & Pretraining	Cropping Dimensions (Pixels)	Tiling Overlap	Loss Function	Optimizer & Initial LR
HED	Edge Det.	VGG-16 (Official checkpoint)	512 × 512	0% (Stride 512)	Multi-side weighted BCE	SGD (lr = 0.01)
RCF	Edge Det.	VGG-16 (Official checkpoint)	512 × 512	0% (Stride 512)	Multi-side class-balanced BCE	SGD (lr = 10^{-6})
D-LinkNet	Edge Det.	ResNet-34 (ImageNet pretrained)	1024 × 1024	0% (Stride 1024)	BCE + Soft Dice Loss	Adam (lr = 10^{-4})
ResUNet-a	Multi-Task	ResNet-50 variant (Scratch)	224 × 224	0% (Stride 224)	Tversky Loss ($\alpha = \beta = 0.5$)	Adam (lr = 10^{-4})
BsiNet	Multi-Task	Custom encoder (Official checkpoint)	512 × 512	0% (Stride 512)	Foreground/Contour CE (NLL) + Distance MSE	Adam (lr = 10^{-4})
TFNet	Multi-Task	FastSAM-x encoder (FastSAM-x pretrained)	512 × 512	0% (Stride 512)	BCE + class-balanced boundary BCE + cosine similarity loss	SGD (lr = 0.1)
E2EC	Instance	DeepSnake Backbone (COCO pretrained)	640 × 640	0% (Stride 640)	Modified Focal + multi-stage Smooth L1	Adam (lr = 10^{-4})
SOLOv2	Instance	ResNet-50-FPN (COCO pretrained)	1024 × 1024	0% (Stride 1024)	Sigmoid Focal (Category) + Dice (Mask)	Adam (lr = 10^{-4})
YOLOv8-seg	Instance	YOLOv8n-seg (COCO Pretrained)	640 × 640	0% (Stride 640)	IoU-based Box Loss + BCE (Mask/Class) + DFL	Adam (lr = 10^{-4})

In the ensuing stages of visualization of model probability maps and quantitative accuracy evaluation, all model outputs were binarized to delineate continuous parcel boundaries. Cropland parcel boundary extraction is inherently subject to extreme class imbalance, as boundary pixels typically make up less than 3% of the total image area. Applying the standard 0.5 probability threshold would result in severe boundary fragmentation and false-negative omissions. For this reason, in line with established benchmark conventions for boundary and edge detection (such as the standard evaluation protocols of HED and RCF), a uniform empirical threshold of T = 0.3 was applied across all model architectures. This operational cutoff maintains boundary recall and topological continuity, and enables an impartial comparison that reflects raw feature discriminability free from model-specific post hoc tuning bias.

2.7. Model Evaluation Metrics

To quantitatively evaluate the performance of different deep learning models for cropland parcel boundary extraction, four commonly used metrics were adopted: precision (P), recall (R), IoU, and F1 score. Because the evaluated models produce different types of outputs, including region masks and explicit edge maps, all predictions were first converted into a unified pixel-level boundary representation before metric calculation.

For models generating parcel masks, binary predictions were first converted into one-pixel-wide external contours via contour extraction. Specifically, external contours were detected and rendered at a one-pixel thickness. For models directly outputting edge

maps, the predicted binary edge maps were adopted without extra contour extraction. In both scenarios, the resulting boundary maps were then thinned into one-pixel skeletons to mitigate discrepancies arising from variable thickness in raw predicted boundaries.

After thinning, both the predicted boundary and the reference boundary were dilated using a 7 × 7 square structuring element. This operation corresponds to a three-pixel tolerance radius on each side of the one-pixel skeleton. Therefore, for imagery with a ground sampling distance of r, the nominal tolerance radius is approximately 3r in map units. This tolerance-based processing reduces excessive penalization caused by minor pixel-level misalignment between predicted and reference parcel boundaries.

In boundary extraction tasks, pixel-wise correspondence is practically unattainable due to manual annotation jitter and rasterization aliasing. It is a well-established convention in the field to define tolerance buffers by pixel multiples (i.e., k pixels, equivalent to k × ground sampling distance (GSD) in ground distance) for multi-resolution comparisons and cross-dataset model evaluations. This approach preserves the relative consistency of evaluation scales across different sensor resolutions [63]. Accordingly, this study implements a scale-aware, symmetric buffer-matching protocol, uniformly configuring the tolerance buffer width to three times the spatial resolution of the respective imagery (R = 3 × GSD, equivalent to 3 pixels). This 3-pixel relaxation parameter is essential to absorb human delineation variance (1–2 pixels), rasterization staircase discretization, and sensor point spread function (PSF) blur across varying spatial resolutions.

As illustrated in the symmetric evaluation scheme in Figure 5, the metrics are computed based on boundary lengths across two complementary matching directions:

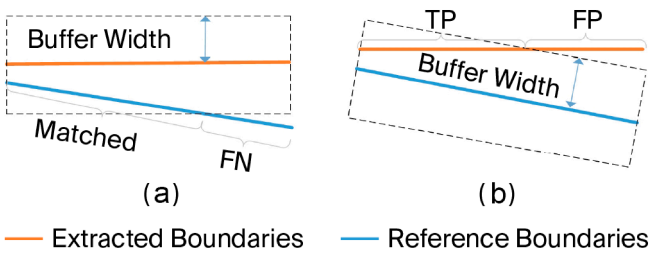


Figure 5. Schematic diagram for the evaluation of cropland parcel boundary extraction results: (**a**) matching of extraction results; (**b**) matching of ground truth references.

Precision Evaluation (Figure 5b): A buffer of radius R is generated around the reference (ground-truth) boundaries.

True Positive (TP): The pixel count of extracted boundaries falling within the reference buffer, indicating correctly delineated boundary segments:

$$TP = |\{(x,y) \in P | Dist((x,y),G) \leq R\}|$$

False Positive (FP): The pixel count of extracted boundaries located outside the reference buffer, representing false alarms or over-segmented noise:

$$FP = |P| - TP$$

Recall Evaluation (Figure 5a): Symmetrically, a buffer of radius R is generated around the extracted (predicted) boundaries.

Matched Reference (Matched): The pixel count of ground-truth boundaries covered by the prediction buffer, representing successfully detected reference segments.

False Negative (FN): The pixel count of ground-truth boundaries remaining outside the prediction buffer, representing missed or omitted boundary segments:

$$FN = |G| - |\{(x,y) \in G | Dist((x,y),P) \leq R\}|$$

This dual-directional distance evaluation establishes a symmetric matching process and inherently prevents double counting by evaluating precision over the extracted prediction set and recall over the ground-truth reference set.

Based on these fundamental elements, the evaluation metrics are mathematically defined as follows:

Precision: This measures the correctness of extracted parcel boundaries. A higher precision value indicates a greater proportion of correctly extracted boundaries and fewer incorrectly identified boundary segments.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{1}$$

Recall: This measures the completeness of identified parcel boundaries. A higher recall value indicates more complete boundary extraction with fewer missed segments.

$$\text{Recall} = \frac{\text{Matched}}{\text{Matched} + \text{FN}} \tag{2}$$

F1 Score: This is the harmonic mean of precision and recall and provides a balanced evaluation of model performance.

$$\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3}$$

IoU: This evaluates the overall accuracy of parcel boundary extraction. A higher IoU value indicates more accurate extraction with fewer incorrect or missed boundary segments.

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FN} + \text{FP}} \tag{4}$$

3. Results

3.1. Public Dataset Comparison

To objectively substantiate the characteristics and contributions of this dataset, Table 4 presents a systematic, multi-dimensional comparison between our benchmark and representative publicly available cropland datasets, including Agriculture-Vision, PASTIS, and AI4Boundaries. First, compared with similar public benchmarks, our benchmark dataset spans diverse landscapes across Europe and Central China, providing topographic heterogeneity across plains, hills, and mountains. Second, in terms of multi-sensor and multi-resolution coverage, our benchmark dataset integrates Global Mapper, GE, Sentinel-2, Pleiades, and UAV photogrammetry sources across an extensive spatial resolution gradient (0.2 m to 10 m). Most importantly, regarding label fidelity, while existing datasets commonly rely on unrectified or directly rasterized cadastral vectors, our benchmark uses an expertly refined workflow to correct temporal drift and digitizing errors systematically.

Table 4. Structural and attribute comparison between the proposed benchmark dataset and representative open cropland datasets.

Dataset	Region	Unique Parcels	Sensors	Spatial Resolution	Spectral/Modality	Annotation Protocol	Topographic Diversity
Agriculture-Vision	United States	3432 parcels	Aerial sensor	0.1–0.2 m	RGB + NIR	Semantic masks (field agricultural anomalies)	Low (mostly flat)
PASTIS	France (4 zones)	~124 k parcels	Sentinel-2	10–20 m	Multispectral + time-series	Parcel masks (crop types)	Low (lowlands/rolling)

Table 4. *Cont*.

Dataset	Region	Unique Parcels	Sensors	Spatial Resolution	Spectral/Modality	Annotation Protocol	Topographic Diversity
AI4Boundaries	7 EU countries	~2.4 M parcels	Sentinel-2, Aerial photos	All areas: 10 m/Partial areas: 1 m	RGB + NIR + NDVI (Normalized Difference Vegetation Index)	Parcel masks (cropland parcel boundary)	Moderate (temperate Europe)
Our Benchmark Dataset	Three Europe & Central China	>180 k unique (>280 k multi-res obs)	Global Mapper, Geospatial Earth, Sentinel-2, Pleiades, UAV photogrammetry	0.2 m/0.5 m/0.6 m/2.4 m/4.8 m/10 m	RGB	Parcel masks (expertly refined parcel boundary)	High (plains, hills, mountains)

3.2. Evaluation of the Extraction Effects of Nine Deep Learning Models

Under uniform sub-meter high-resolution conditions, this study comparatively evaluated the boundary extraction performance and architectural capabilities of nine deep learning models across heterogeneous agricultural landscapes, with the detailed evaluation metrics summarized in Table 5.

Table 5. Performance comparison of nine deep learning methods across six study areas.

Area	Evaluation Metrics	HED	RCF	D-LinkNet	BsiNet	ResUNet-a	TFNet	E2EC	SOLOv2	YOLOv8-Seg
Denmark	Precision (%)	31.00	30.31	44.75	57.46	**72.53**	55.19	35.64	55.56	48.81
	Recall (%)	44.25	61.09	37.28	**62.46**	57.62	58.52	40.34	50.55	53.58
	F1 Score (%)	33.68	37.76	39.89	58.81	**61.65**	56.80	34.88	49.99	50.03
	IoU (%)	21.66	24.66	25.48	44.49	**47.14**	39.67	21.53	35.81	34.63
Netherlands	Precision (%)	63.24	43.83	70.30	78.76	**79.36**	78.01	57.75	56.50	50.95
	Recall (%)	67.81	73.73	74.34	**80.27**	68.70	68.87	68.03	45.86	56.15
	F1 Score (%)	62.30	51.56	71.60	**79.12**	70.72	73.16	61.18	49.29	52.99
	IoU (%)	48.09	37.14	56.42	**69.03**	58.43	57.68	44.97	33.26	37.47
Spain	Precision (%)	58.11	50.21	63.10	69.33	**76.94**	70.29	57.95	62.69	61.70
	Recall (%)	59.91	65.43	57.57	75.15	**76.93**	53.71	53.12	43.49	61.33
	F1 Score (%)	57.49	55.51	60.06	70.83	**75.53**	60.89	54.94	49.80	61.01
	IoU (%)	41.48	39.50	43.16	56.67	**62.80**	43.77	38.54	33.75	44.53
Jiangxia District	Precision (%)	66.15	62.14	73.70	**89.84**	66.76	69.84	58.13	50.12	63.30
	Recall (%)	50.68	61.95	75.66	**84.29**	61.47	67.77	64.07	46.52	57.63
	F1 Score (%)	55.60	60.80	74.64	**86.58**	62.17	68.79	60.31	46.34	60.04
	IoU (%)	39.32	44.41	59.78	**77.92**	47.82	52.43	44.71	31.26	43.84
Echeng District	Precision (%)	61.09	53.79	65.47	**84.43**	54.06	44.73	48.74	42.38	56.67
	Recall (%)	61.18	68.06	54.69	**69.13**	47.16	41.61	52.58	47.75	45.71
	F1 Score (%)	60.67	59.65	59.42	**75.06**	47.55	43.11	47.71	40.18	49.75
	IoU (%)	44.54	43.30	42.83	**64.04**	33.11	27.48	32.00	26.21	33.66
Tongshan County	Precision (%)	**78.63**	73.58	67.74	41.06	57.54	69.59	63.68	61.90	64.61
	Recall (%)	57.94	68.91	**70.30**	36.22	55.67	68.36	61.65	41.02	20.97
	F1 Score (%)	66.09	**70.85**	68.81	36.56	52.83	68.97	60.89	47.30	30.85
	IoU (%)	49.92	**55.17**	52.58	23.31	37.59	52.63	44.45	31.59	18.58

Note: Bold values indicate the best performance.

Table 5 presents a comprehensive performance comparison of the nine deep learning models across the six study areas, revealing pronounced disparities across model paradigms, individual architectures, and regional landscapes.

Regarding cross-regional landscape divergence, model efficacy exhibited geographic clustering. In plain and hilly regions, multi-task semantic segmentation architectures (BsiNet, ResUNet-a, and TFNet) generally achieved favorable boundary delineation performance. Specifically, BsiNet yielded the highest overall metrics in Jiangxia District (fragmented plains), Echeng District (rolling hills), and the Netherlands (large-scale, regular rectangular parcels). In Jiangxia District, for example, its precision, recall, F1 score, and IoU reached 89.84%, 84.29%, 86.58%, and 77.92%, respectively. Meanwhile, across landscapes featuring non-orthogonal parcel orientations and irregular polygonal contours, such as Denmark and Spain, ResUNet-a achieved F1 scores of 61.65% and 75.53%, respectively. In contrast, in the steep mountainous terrain of Tongshan County, where pronounced topographic relief is coupled with irregular, meandering parcel contours, model performance yields results distinct from those in the plains and hilly areas. Multi-task frameworks that performed robustly in low-relief terrain experienced a marked accuracy decline, with the F1 score of BsiNet dropping to 36.56%. Conversely, local edge detection networks (HED, RCF, and D-LinkNet) exhibited greater resilience in complex mountainous terrain. HED attained the highest precision (78.63%), D-LinkNet achieved the highest recall (70.30%), and RCF produced the top F1 score (70.85%) and boundary IoU (55.17%). The three instance segmentation models (E2EC, SOLOv2, and YOLOv8-seg) achieved relatively low overall extraction accuracy.

Cross-paradigm comparisons reflect behavioral differences aligned with distinct architectural designs. Within the multi-task semantic segmentation paradigm, clear specialization emerged among architectures. BsiNet excels in regular contiguous plains due to its explicit distance field repulsion, whereas ResUNet-a is better suited to irregular geometric polygons. The TFNet exhibited cross-landscape stability. Benefiting from zero-shot boundary priors provided by the Segment Anything Model (SAM) and decoupled two-flow interactions, TFNet achieved an F1 score of 68.97% and an IoU of 52.63% in mountainous Tongshan County. This substantially outperformed BsiNet (36.56%) and ResUNet-a (52.83%), suggesting a potential benefit from foundation-model-derived priors in handling cast shadows and low-contrast boundaries. Instance segmentation models (E2EC, SOLOv2, and YOLOv8-seg) consistently yielded lower pixel-level boundary extraction metrics across all study sites, with F1 scores predominantly situated in the 40% to 61% range. This trend is consistent with spatial upsampling smoothing inherent to mask prediction from deep feature maps under strict buffer-matching protocols.

Figure 6 shows a visual comparison of the nine models' performance in extracting cropland parcel boundaries in the six study areas. The study areas in Denmark (Figure 6a), the Netherlands (Figure 6b), and Spain (Figure 6c) feature flat terrain, large-scale cropland parcels, and straight, well-defined boundaries. Except for E2EC and SOLOv2, which exhibited significant omissions of target boundaries, the remaining models were able to effectively identify cropland parcel boundaries. Among them, ResUNet-a and BsiNet, leveraging the advantages of multiscale feature fusion, were able to maintain the integrity and continuity of cropland boundaries relatively well; HED and RCF tended to misidentify cropland ridges and surface textures as boundaries in the Danish study area, generating a large amount of false boundary noise; in the Spanish study area, HED, RCF, D-LinkNet, and YOLOv8-seg exhibited significant omissions of weak, low-contrast boundaries between adjacent cropland parcels.

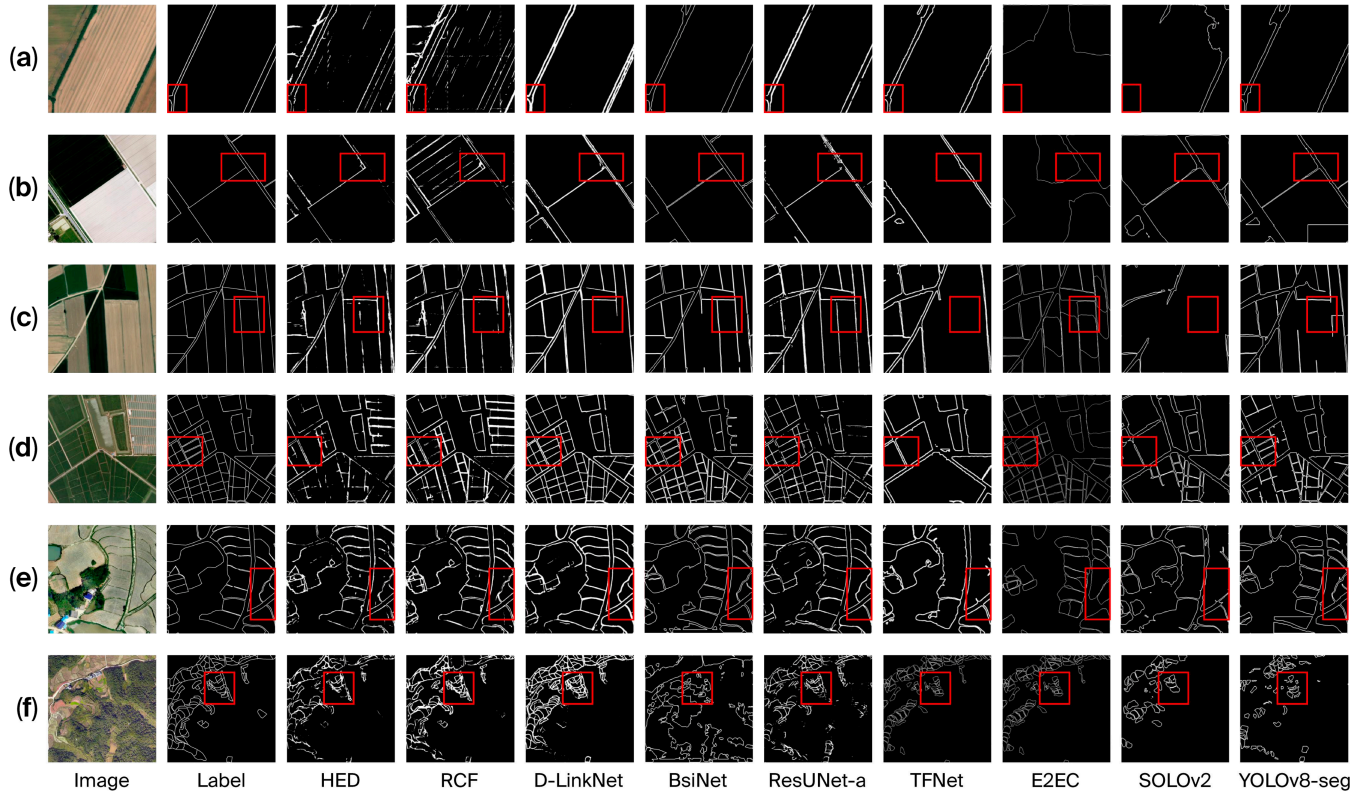


Figure 6. Visualization results of different models. The presented image samples are from (**a**) Denmark, (**b**) Netherlands, (**c**) Spain, (**d**) Jiangxia District, (**e**) Echeng District, and (**f**) Tongshan County.

Jiangxia District (Figure 6d) also features flat terrain, but the parcels are fragmented and densely packed with complex boundaries, placing higher demands on the models' ability to perceive fine details. In this area, BsiNet and D-LinkNet reconstructed fragmented parcel boundaries with relatively good continuity of fragmented parcels, while the extraction results from HED and RCF exhibit obvious boundary discontinuities, and the omission issues in YOLOv8-seg and SOLOv2 are particularly prominent. In Echeng District (Figure 6e), due to undulating terrain, parcel boundaries often take on irregular, curved shapes. In this scenario, BsiNet continues to deliver good extraction results, while HED, RCF, and D-LinkNet are also able to capture parcel boundaries with relative stability; E2EC and SOLOv2 lack sufficient adaptability to curved boundaries, resulting in very obvious boundary breaks and missed detections. Tongshan County (Figure 6f) features mountainous terrain with scattered cropland parcels. Due to interference from the complex topography, the extraction accuracy of all models decreased to varying degrees. In this scenario, traditional edge detection models demonstrated some adaptability; among them, D-LinkNet was able to identify more valid boundary segments, while TFNet and BsiNet's results contained a large number of false positives.

3.3. Effects of Different Spatial Resolutions on Cropland Parcel Boundary Extraction

To evaluate the practical applicability of operational remote sensing data tiers for cropland parcel boundary extraction, this study examined four authentic data sources across a spatial resolution gradient in the Netherlands: Global Mapper aerial orthomosaics (0.6 m), Geospatial Earth satellite imagery (2.4 m), Pleiades satellite imagery (4.8 m), and Sentinel-2 multispectral observations (10 m). Conducting this benchmark within a single geographic region serves to minimize broad landscape heterogeneity, keeping macroscopic attributes such as flat topography, general farming practices, and average

parcel size (consistently around 10 hm^2) relatively uniform. These datasets capture realistic operational tiers accessible in applied agricultural monitoring. Consequently, performance variations across these tiers reflect the combined influence of spatial resolution, native sensor imaging characteristics, and associated operational observational conditions under real-world monitoring workflows. The results are presented in Table 6.

Table 6. Performance comparison of nine deep learning models across four operational spatial resolution tiers in the Netherlands.

Site	Evaluation Metrics	HED	RCF	D-LinkNet	BsiNet	ResUNet-a	TFNet	E2EC	SOLOv2	YOLOv8-Seg
Netherlands (Spatial resolution 0.6 m)	Precision (%)	63.24	43.83	70.30	78.76	**79.36**	78.01	57.75	56.50	50.95
	Recall (%)	67.81	73.73	74.34	**80.27**	68.70	68.87	68.03	45.86	56.15
	F1 Score (%)	62.30	51.56	71.60	**79.12**	70.72	73.16	61.18	49.29	52.99
	IoU (%)	48.09	37.14	56.42	**69.03**	58.43	57.68	44.97	33.26	37.47
Netherlands (Spatial resolution 2.4 m)	Precision (%)	39.38	39.04	39.38	75.96	48.19	**78.15**	47.73	67.75	52.69
	Recall (%)	24.25	31.24	24.25	65.82	50.41	49.54	41.73	**66.65**	30.37
	F1 Score (%)	29.05	33.94	29.05	**69.00**	47.74	60.64	42.92	63.75	36.35
	IoU (%)	17.50	21.00	17.50	**53.77**	31.95	43.52	27.90	48.74	23.63
Netherlands (Spatial resolution 4.8 m)	Precision (%)	76.04	75.77	70.52	80.84	73.36	78.94	69.83	**81.35**	75.22
	Recall (%)	48.62	55.26	64.18	65.44	60.22	60.63	55.08	**72.84**	24.81
	F1 Score (%)	58.57	63.11	66.81	71.69	64.43	68.58	59.17	**76.70**	34.49
	IoU (%)	41.86	46.65	50.37	56.77	48.94	52.18	42.86	**62.28**	22.23
Netherlands (Spatial resolution 10 m)	Precision (%)	23.71	24.00	26.69	59.62	34.15	73.11	48.34	72.77	**74.30**
	Recall (%)	10.01	11.02	15.65	22.08	22.26	21.01	16.75	**48.90**	10.95
	F1 Score (%)	13.29	14.55	19.20	32.06	25.34	32.64	21.69	**58.02**	16.36
	IoU (%)	7.16	7.89	10.65	19.19	14.70	19.50	12.30	**43.55**	10.03

Note: Bold values indicate the best performance.

As shown in Table 6, at the sub-meter resolution of 0.6 m, multi-task semantic segmentation models generally surpassed other paradigms. BsiNet achieved the highest F1 score (79.12%) and boundary IoU (69.03%), followed by D-LinkNet (71.60%) and ResUNet-a (70.72%). The edge detection model HED also exhibited competitive boundary delineation capability, achieving an F1 score of 62.30%. In contrast, the two instance segmentation networks, SOLOv2 and YOLOv8-seg, yielded lower pixel-level boundary extraction accuracy, recording F1 scores of 49.29% and 52.99%, respectively.

When the spatial resolution coarsened to 2.4 m, BsiNet maintained the highest overall performance, retaining an F1 score of 69.00% and an IoU of 53.77%. SOLOv2 ranked second across the benchmark with an F1 score of 63.75%, visibly exceeding ResUNet-a (47.74%) and E2EC (42.92%). At a spatial resolution of 4.8 m, SOLOv2 achieved the highest F1 score (76.70%) and precision (81.35%), suggesting that position-driven instance priors may provide favorable adaptability to this intermediate scale. Concurrently, BsiNet preserved robust detection performance with an F1 score of 71.69%, followed by D-LinkNet (66.81%) and ResUNet-a (64.43%), whereas YOLOv8-seg recorded the lowest F1 score of 34.49%. Under the coarsest resolution of 10 m, the scale resilience of SOLOv2 became even more pronounced, maintaining the highest F1 score (58.02%), IoU (43.55%), and recall (48.90%). In contrast, all remaining architectures experienced severe performance collapses, with their boundary IoU decreasing below 20%.

Except for SOLOv2 and YOLOv8-seg, the remaining seven models exhibited an initial decrease, a subsequent rebound, and then a severe decline, as shown in Figure 7. Among the evaluated models, the edge detection models (HED, RCF, and D-LinkNet) were the most sensitive to resolution changes; the multi-task semantic segmentation models (BsiNet, ResUNet-a, and TFNet) exhibited greater stability in extraction accuracy and were less

susceptible to resolution variations; and the three instance segmentation models (E2EC, SOLOv2, and YOLOv8-seg) differed considerably in how resolution affected their extraction performance, with SOLOv2 performing best overall across resolution changes. From the overall trends of all models across different resolutions, most models showed only small differences in extraction accuracy between 4.8 m medium-resolution imagery and 0.6 m sub-meter imagery.

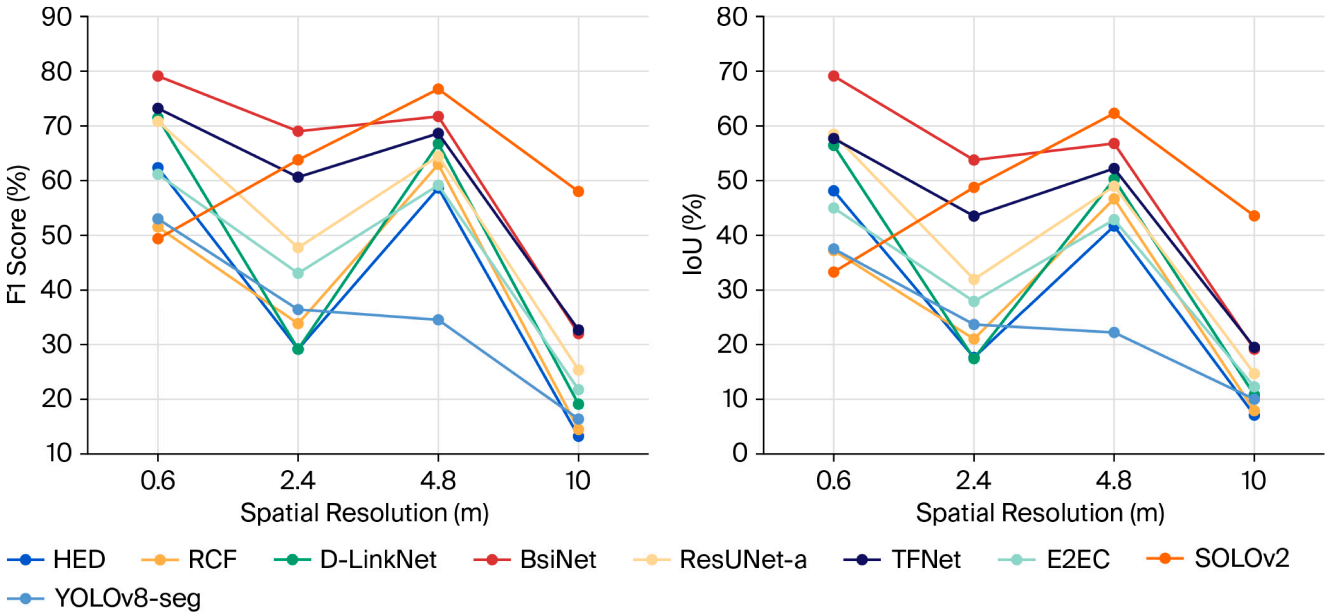


Figure 7. Trend of boundary extraction accuracy across four operational spatial resolution tiers.

3.4. Effects of Parcel Scale on Cropland Parcel Boundary Extraction

To evaluate the influence of parcel scale on boundary extraction accuracy while minimizing severe terrain interference, this study examined four agricultural areas representing an empirical parcel scale gradient: Jiangxia District (0.22 hm^2), Spain (1.5 hm^2), Denmark (6.0 hm^2), and the Netherlands (10.0 hm^2). Table 7 shows the detailed results.

Table 7. Performance comparison of nine deep learning models across four parcel-scale gradients.

Site	Evaluation Metrics	HED	RCF	D-LinkNet	BsiNet	ResUNet-a	TFNet	E2EC	SOLOv2	YOLOv8-Seg
Jiangxia District (Parcel scale 0.22 hm^2)	Precision (%)	66.15	62.14	73.70	**89.84**	66.76	69.84	58.13	50.12	63.30
	Recall (%)	50.68	61.95	75.66	**84.29**	61.47	67.77	64.07	46.52	57.63
	F1 Score (%)	55.60	60.80	74.64	**86.58**	62.17	68.79	60.31	46.34	60.04
	IoU (%)	39.32	44.41	59.78	**77.92**	47.82	52.43	44.71	31.26	43.84
Spain (Parcel scale 1.5 hm^2)	Precision (%)	58.11	50.21	63.10	69.33	**76.94**	70.29	57.95	62.69	61.70
	Recall (%)	59.91	65.43	57.57	75.15	**76.93**	53.71	53.12	43.49	61.33
	F1 Score (%)	57.49	55.51	60.06	70.83	**75.53**	60.89	54.94	49.80	61.01
	IoU (%)	41.48	39.50	43.16	56.67	**62.80**	43.77	38.54	33.75	44.53
Denmark (Parcel scale 6 hm^2)	Precision (%)	31.00	30.31	44.75	57.46	**72.53**	55.19	35.64	55.56	48.81
	Recall (%)	44.25	61.09	37.28	62.46	**57.62**	58.52	40.34	50.55	53.58
	F1 Score (%)	33.68	37.76	39.89	58.81	**61.65**	56.80	34.88	49.99	50.03
	IoU (%)	21.66	24.66	25.48	44.49	**47.14**	39.67	21.53	35.81	34.63
Netherlands (Parcel scale 10 hm^2)	Precision (%)	63.24	43.83	70.30	78.76	**79.36**	78.01	57.75	56.50	50.95
	Recall (%)	67.81	73.73	74.34	**80.27**	68.70	68.87	68.03	45.86	56.15
	F1 Score (%)	62.30	51.56	71.60	**79.12**	70.72	73.16	61.18	49.29	52.99
	IoU (%)	48.09	37.14	56.42	**69.03**	58.43	57.68	44.97	33.26	37.47

Note: Bold values indicate the best performance.

The results in Table 7 reveal that the two multi-task semantic segmentation architectures, BsiNet and ResUNet-a, generally achieved superior boundary delineation performance across the evaluated scale tiers. Specifically, BsiNet attained the highest accuracy at both the small-scale and large-scale extremes. In Jiangxia District (0.22 hm^2), BsiNet led across all evaluated metrics, recording a precision of 89.84%, recall of 84.29%, boundary IoU of 77.92%, and an F1 score of 86.58%. Similarly, at the 10.0 hm^2 scale in the Netherlands, BsiNet achieved the top ranking with a recall of 80.27%, an F1 score of 79.12%, and an IoU of 69.03%. Conversely, ResUNet-a exhibited favorable adaptability at intermediate scale tiers, leading across all evaluation metrics in Spain (1.5 hm^2, F1 score: 75.53%) and Denmark (6.0 hm^2, F1 score: 61.65%). TFNet maintained consistently high performance across all four parcel scale gradients, with F1 scores ranging from 56.80% to 73.16%. D-LinkNet also maintained competitiveness in Jiangxia District and the Netherlands, achieving F1 scores of 74.64% and 71.60%, respectively. By contrast, edge detection models (HED and RCF) and instance segmentation models (E2EC, SOLOv2, YOLOv8-seg) generally yielded lower boundary extraction metrics.

Figure 8 depicts the performance trajectories of all evaluated models across the parcel scale spectrum. When the parcel scale increased from 0.22 hm^2 to 1.5 hm^2, the performance changes varied among models, with some showing increases and others showing decreases, as indicated in Figure 8. When the parcel scale increased from 1.5 hm^2 to 6.0 hm^2, all models except SOLOv2 exhibited a declining trend in both F1 score and IoU. When the parcel scale increased to 10.0 hm^2, compared with the 6.0 hm^2 scale, all models except SOLOv2 exhibited an upward trend in both F1 score and IoU. The edge detection models showed relatively larger improvements. SOLOv2 slightly decreased during this stage. Across the entire scale continuum, the multi-task semantic segmentation models generally occupied the upper range of the performance trajectories, demonstrating structural robustness.

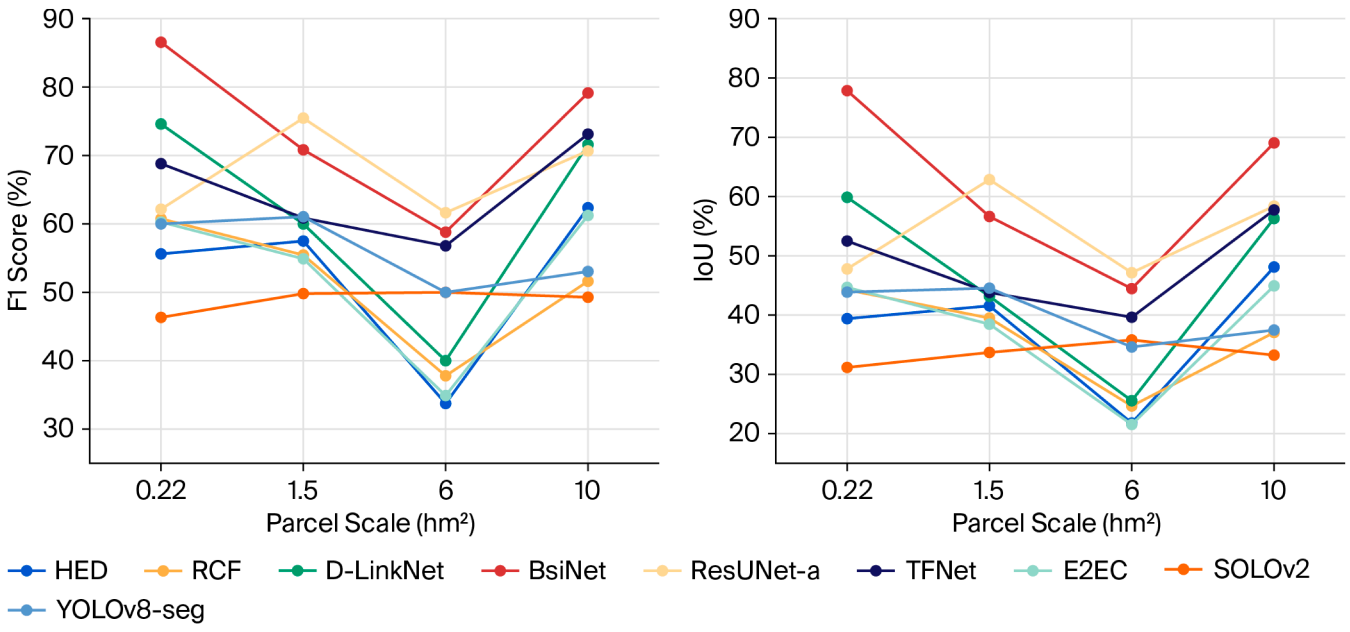


Figure 8. Trend of boundary extraction accuracy across four parcel scale gradients.

3.5. Effects of Terrain Conditions on Cropland Parcel Boundary Extraction

The three Central China study areas constitute a topographic gradient transect with quantified relief amplitudes: the Jiangxia plain (56 m), Echeng hills (470 m), and Tongshan mountainous area (1637 m) (Table 1). Instead of evaluating static performance rankings, this analysis focuses on cross-terrain sensitivity dynamics (Table 8).

Table 8. Performance comparison of nine deep learning models across three terrains.

Site	Evaluation Metrics	HED	RCF	D-LinkNet	BsiNet	ResUNet-a	TFNet	E2EC	SOLOv2	YOLOv8-Seg
Jiangxia District (Plain)	Precision (%)	66.15	62.14	73.70	**89.84**	66.76	69.84	58.13	50.12	63.30
	Recall (%)	50.68	61.95	75.66	**84.29**	61.47	67.77	64.07	46.52	57.63
	F1 Score (%)	55.60	60.80	74.64	**86.58**	62.17	68.79	60.31	46.34	60.04
	IoU (%)	39.32	44.41	59.78	**77.92**	47.82	52.43	44.71	31.26	43.84
Echeng District (Hill)	Precision (%)	61.09	53.79	65.47	**84.43**	54.06	44.73	48.74	42.38	56.67
	Recall (%)	61.18	68.06	54.69	**69.13**	47.16	41.61	52.58	47.75	45.71
	F1 Score (%)	60.67	59.65	59.42	**75.06**	47.55	43.11	47.71	40.18	49.75
	IoU (%)	44.54	43.30	42.83	**64.04**	33.11	27.48	32.00	26.21	33.66
Tongshan County (Mountain)	Precision (%)	**78.63**	73.58	67.74	41.06	57.54	69.59	63.68	61.90	64.61
	Recall (%)	57.94	68.91	**70.30**	36.22	55.67	68.36	61.65	41.02	20.97
	F1 Score (%)	66.09	**70.85**	68.81	36.56	52.83	68.97	60.89	47.30	30.85
	IoU (%)	49.92	**55.17**	52.58	23.31	37.59	52.63	44.45	31.59	18.58

Note: Bold values indicate the best performance.

Along this topographic gradient, model architectures exhibited substantially divergent performance degradation trajectories. Multi-task semantic segmentation architectures show contrasting resilience across relief tiers. BsiNet displayed sensitivity to rising topographic amplitude. Its F1 score decreased by 11.52% from plain to hilly terrain, followed by a further decrease of 38.50% when relief reaches 1637 m in Tongshan County. This vulnerability may plausibly arise from degraded continuous Euclidean distance field representations over fractured, shadow-obscured slope geometries. In comparison, the foundation model-guided TFNet showed promising stability across varying terrain. It achieved competitive performance on the plain (F1 score: 68.79%) and retained an F1 score of 68.97% and boundary IoU of 52.63% in steep mountainous terrain, considerably outperforming BsiNet (36.56%) and ResUNet-a (52.83%). This contrast suggests that the integration of SAM zero-shot boundary priors and decoupled two-flow representations may help mitigate the topological deterioration commonly observed in conventional multi-task architectures under mountain shadowing.

Figure 9 further summarizes performance trajectories of all evaluated models along this topographic gradient. BsiNet and YOLOv8-seg followed a monotonic downward trajectory across the terrain gradient. In contrast, HED, RCF, and D-LinkNet exhibited upward or rebounding trajectories. ResUNet-a, E2EC, and TFNet traced a non-linear, dip-and-rebound pattern characterized by lower performance in hilly terrain followed by recovery in mountainous terrain, with TFNet reaching metrics comparable to RCF and D-LinkNet in the mountain setting. Collectively, these trajectories suggest an apparent cross-over pattern between plain and mountainous landscapes, where models that led in the plain setting experienced steep declines, whereas models with lower plain metrics showed marked upward or rebounding trends in mountainous terrain.

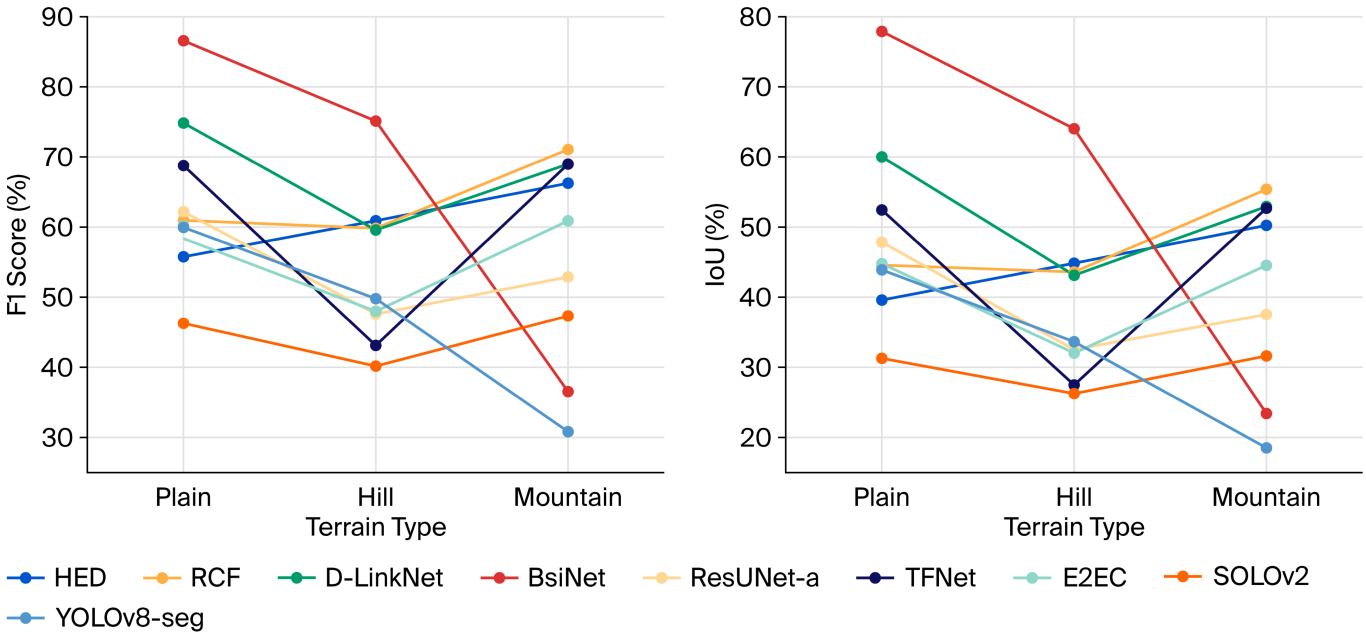


Figure 9. Trend of boundary extraction accuracy across three terrains.

4. Discussion

Public cropland datasets such as PASTIS and AI4Boundaries commonly rely on administrative cadastral records derived from farmer declarations. These records frequently contain label noise, such as omitted boundary paths, drainage ditches, and geometric misalignments, which can bias benchmark evaluation metrics. To reduce label errors, this study used an expertly refined workflow to inspect and rectify cadastral vectors against contemporaneous high-resolution imagery. This procedure aligns reference boundaries with the visible parcel boundaries in imagery acquired at the corresponding date, mitigating label discrepancies caused by temporal lags and land management changes such as field consolidation and subdivision. Furthermore, the resulting dataset covers diverse agricultural landscapes, spanning large, regular parcels in Western Europe (averaging up to 10.0 hm^2) and fragmented smallholder fields in Central China (averaging 0.05 hm^2), with imagery resolutions ranging from 0.2 m to 10 m.

Grounded in our constructed benchmark dataset, although sensor platforms, data volumes, and topographic conditions naturally co-vary across the study areas [64], this work offers an empirical baseline reflecting the real-world observational variability encountered in operational agricultural monitoring. Nevertheless, several inherent limitations must be acknowledged when interpreting these findings and guide future research. Geographic coverage and intra-regional sampling density of the current dataset remain bounded. In particular, the dataset does not yet encompass agricultural scenarios with significant heterogeneity, such as the US Midwest and North China Plain, tropical perennial plantations in Southeast Asia and Brazil, complex flooded paddy terraces, and rainfed smallholder mosaics in sub-Saharan Africa). Future work will expand the dataset in three dimensions, geographic coverage, crop varieties, and agricultural management systems, to enhance and refine the data selection strategies offered. The current evaluation framework focuses primarily on in-domain spatial adaptability rather than cross-domain zero-shot generalization. Models were benchmarked within regional subsets to isolate architecture-specific responses to local field configurations. However, direct cross-continental transferability was not systematically quantified. We acknowledge the theoretical continuum spanning from localized spatial block partitioning to systematic zero-shot Leave-One-Region-Out

(LORO) cross-domain transfer as a vital direction. Future work will benchmark transfer learning, domain adaptation, and few-shot fine-tuning protocols to establish standardized cross-domain transfer baselines.

Models exhibit a non-monotonic performance trajectory across spatial resolutions. In large-scale agricultural plains (10.0 hm^2), model accuracy did not degrade continuously from 0.6 m to 10 m; instead, performance dipped at 2.4 m before rebounding at 4.8 m to levels comparable to 0.6 m, followed by a sharp drop at 10 m. At the 2.4 m resolution, models may encounter an intermediate scale trade-off. Pure boundary pixels are reduced, yet high-frequency intra-parcel textures (such as tractor tramlines and furrow shadows) remain visible and generate spurious edge responses. Transitioning to 4.8 m may function as a spatial low-pass-like filter that attenuates intra-field textural noise while preserving prominent inter-parcel boundaries, with broader receptive fields capturing complete macro-topologies. When resolution degraded to 10 m, pixel dimensions exceed physical boundary widths, which may lead to mixed-pixel deterioration. Under this extreme degradation, position-driven models (SOLOv2) maintained higher contour closure probably because grid-based center priors rely less heavily on degraded boundary gradients. Because these resolution tiers were compiled from authentic operational data sources (Global Mapper aerial orthomosaics, Geospatial Earth, Pleiades, and Sentinel-2), spatial resolution inherently co-varies with native sensor characteristics, including modulation transfer functions (MTF), signal-to-noise ratios, radiometric calibrations, and minor acquisition timing differences. Consequently, the observed performance disparities capture the composite influence of real-world operational imaging tiers rather than GSD in pure physical isolation. As a next step, we will further investigate the influence of spatial resolution by constructing a benchmark from a single ultra-high-resolution source scene with an identical spatial footprint and deriving multiple resolution tiers through controlled spatial downsampling and realistic PSF/MTF filtering.

Multi-task semantic segmentation models such as BsiNet and ResUNet-a leverage a joint multi-task and distance field supervision architecture to adapt to the characteristics of contiguous cropland. Within densely contiguous regions of remote sensing imagery, the numerical gradient generated by the distance field may provide strong geometric constraints. This boundary-equivalent repulsion effect plausibly helps alleviate the prevalent issue of parcel boundary adhesion in contiguous imagery. When integrated with a spatial grouping enhancement module, such architectures appear better suited to capturing the textural features of fragmented parcels, which may facilitate more refined delineation of small-scale parcel boundaries. Because the evaluated scale gradient (spanning 0.22 hm^2 to 10.0 hm^2) was assembled across distinct geographic landscapes, parcel area inherently co-varies with geometric regularity (such as compactness and aspect ratio), boundary spatial orientation, landscape fragmentation, and localized spectral contrast. Consequently, the observed performance disparities across scale tiers cannot be attributed solely to parcel dimensions in pure isolation. To decouple parcel scale from these confounding geometric and environmental factors, in our future work, we intend to leverage large-scale cadastral databases to conduct stratified quantile sampling, holding geometric compactness and perimeter-to-area ratios constant across varying area tiers, while employing generalized additive models (GAM) to statistically isolate and quantify the independent contribution of parcel area to boundary extraction accuracy.

Multi-task semantic segmentation models (BsiNet and ResUNet-a), the foundation model-guided TFNet, and edge detectors (HED and RCF) exhibited distinct architectural inductive biases across topographic conditions. In low-relief, contiguous plains, multi-task models leveraged distance field supervision, where numerical gradients directed toward parcel margins generate a boundary repulsion-like effect that helps alleviate parcel adhe-

sion. In steep mountainous terrain, however, severe terrain undulation and cast shadows may disrupt image-wide spatial continuity, which may degrade feature representations in architectures relying on continuous global semantic masks and distance fields. Conversely, edge detection models such as HED and RCF relied on multiscale side outputs and localized receptive fields. Although cast shadows attenuated broader contextual information, high-frequency spectral transitions along land cover margins (such as cropland/forest interfaces) often remained locally detectable, providing a plausible explanation for why edge detectors retained higher topological stability in rugged terrain. We acknowledge that while the hilly and mountainous sites (Echeng District and Tongshan County) utilized identical low-altitude UAV photogrammetric workflows to establish a relatively comparable empirical baseline, our terrain evaluation remains an observational comparison rather than a fully decoupled single-factor physical experiment. In future work, we will investigate physical topographic radiometric normalization (such as C-correction) alongside multimodal three-dimensional data fusion (coupling optical imagery with SAR [65] and LiDAR [66]) to actively compensate for the geometric distortions and shadow occlusions inherent to steep mountainous environments.

5. Conclusions

In this study, we constructed a standardized, multi-source cropland parcel boundary benchmark dataset encompassing over 180,000 expert-rectified unique physical parcels across diverse agricultural landscapes in Europe and Central China. Evaluation of nine representative deep learning models across observational and environmental gradients yielded three conclusions:

(1) Spatial resolution exerts a non-linear influence on boundary extraction fidelity, where some medium-resolution data conditions (such as 4.8 m imagery in large-scale fields) retained competitive boundary extraction performance for certain models.

(2) Empirical evaluations across four regional parcel scales (0.22 hm^2 in Jiangxia, 1.5 hm^2 in Spain, 6.0 hm^2 in Denmark, and 10.0 hm^2 in the Netherlands) indicate that boundary extraction accuracy does not exhibit a simple positive correlation with parcel scale.

(3) Terrain complexity is one of the primary factors closely associated with the accuracy drop in mountainous cropland parcel extraction. Multi-task semantic segmentation models excel in plains and hills but degrade sharply in rugged mountainous terrain, where local gradient-driven edge detection models exhibit superior topological resilience.

Overall, this benchmark provides an open dataset and comparative empirical observations across multiple agricultural and topographic settings. These results offer a preliminary baseline that may assist future research and model selection in cropland parcel boundary extraction.

Author Contributions: Conceptualization, Z.L. and P.C.; methodology, H.M.; software, J.H.; validation, B.G. and M.Z.; formal analysis, W.Y.; investigation, H.M.; resources, J.H.; data curation, P.C.; writing—original draft preparation, H.M.; writing—review and editing, P.C.; visualization, J.H.; supervision, J.H.; project administration, B.X.; funding acquisition, Z.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Foundation of Key Laboratory of Fertilization from Agricultural Wastes, Ministry of Agriculture and Rural Affairs (No. KLFAW202503), Foundation of Key Laboratory of Huang-Huai-Hai Smart Agricultural Technology (No. 202502), Open Project Program of Fujian Key Laboratory of Big Data Application and Intellectualization for Tea Industry, Wuyi University (No. FKLBDAITI202310), and National Natural Science Foundation of China (No. 42201397).

Data Availability Statement: The original data presented in this study are openly available in Science Data Bank at https://doi.org/10.57760/sciencedb.16930.

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

Abbreviations

NIR	Near-Infrared
MTF	Modulation Transfer Function
PASTIS	Panoptic Agricultural Satellite Time Series
SAM	Segment Anything Model
BCE	Binary Cross-Entropy
CE	Cross-Entropy
NLL	Negative Log-Likelihood
MSE	Mean Squared Error
DFL	Distribution Focal Loss
GSD	Ground Sampling Distance
PSF	Point Spread Function

References

1. Waldner, F.; Diakogiannis, F.I. Deep learning on edge: Extracting field boundaries from satellite images with a convolutional neural network. *Remote Sens. Environ.* **2020**, *245*, 111741. [CrossRef]
2. Kerner, H.; Chaudhari, S.; Ghosh, A.; Robinson, C.; Ahmad, A.; Choi, E.; Jacobs, N.; Holmes, C.; Mohr, M.; Dodhia, R.; et al. Fields of The World: A machine learning benchmark dataset for global agricultural field boundary segmentation. In Proceedings of the Thirty-Ninth AAAI Conference on Artificial Intelligence (AAAI-25), Philadelphia, PA, USA, 25 February–4 March 2025; pp. 28151–28159. [CrossRef]
3. d'Andrimont, R.; Verhegghen, A.; Lemoine, G.; Kempeneers, P.; Meroni, M.; van der Velde, M. From parcel to continental scale—A first European crop type map based on Sentinel-1 and LUCAS Copernicus in-situ observations. *Remote Sens. Environ.* **2021**, *266*, 112708. [CrossRef]
4. Fetai, B.; Grigillo, D.; Lisec, A. Revising Cadastral Data on Land Boundaries Using Deep Learning in Image-Based Mapping. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 298. [CrossRef]
5. Battude, M.; Al Bitar, A.; Morin, D.; Cros, J.; Huc, M.; Marais Sicre, C.; Le Dantec, V.; Demarez, V. Estimating maize biomass and yield over large areas using high spatial and temporal resolution Sentinel-2 like remote sensing data. *Remote Sens. Environ.* **2016**, *184*, 668–681. [CrossRef]
6. Schneider, M.; Schelte, T.; Schmitz, F.; Körner, M. EuroCrops: The largest harmonized open crop dataset across the European Union. *Sci. Data* **2023**, *10*, 612. [CrossRef]
7. Suarez, L.A.; Robson, A.; Brinkhoff, J. Early-Season forecasting of citrus block-yield using time series remote sensing and machine learning: A case study in Australian orchards. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *122*, 103434. [CrossRef]
8. Barriere, V.; Claverie, M.; Schneider, M.; Lemoine, G.; d'Andrimont, R. Boosting crop classification by hierarchically fusing satellite, rotational, and contextual data. *Remote Sens. Environ.* **2024**, *305*, 114110. [CrossRef]
9. Xiao, W.; Xu, S.; He, T. Mapping Paddy Rice with Sentinel-1/2 and Phenology-, Object-Based Algorithm—A Implementation in Hangjiahu Plain in China Using GEE Platform. *Remote Sens.* **2021**, *13*, 990. [CrossRef]
10. Gomez Selvaraj, M.; Vergara, A.; Montenegro, F.; Alonso Ruiz, H.; Safari, N.; Raymaekers, D.; Ocimati, W.; Ntamwira, J.; Tits, L.; Omondi, A.B.; et al. Detection of banana plants and their major diseases through aerial images and machine learning methods: A case study in DR Congo and Republic of Benin. *ISPRS J. Photogramm. Remote Sens.* **2020**, *169*, 110–124. [CrossRef]
11. Radočaj, D.; Plaščak, I.; Jurišić, M. Global navigation satellite systems as state-of-the-art solutions in precision agriculture: A review of studies indexed in the Web of Science. *Agriculture* **2023**, *13*, 1417. [CrossRef]
12. Yao, Z.; Zhao, C.; Zhang, T. Agricultural machinery automatic navigation technology. *iScience* **2024**, *27*, 108714. [CrossRef]
13. Birinyi, E.; Kristóf, D.; Hollós, R.; Barcza, Z.; Kern, A. Large-scale maize condition mapping to support agricultural risk management. *Remote Sens.* **2024**, *16*, 4672. [CrossRef]
14. Koeva, M.; Bennett, R.; Persello, C. Remote Sensing for Land Administration 2.0. *Remote Sens.* **2022**, *14*, 4359. [CrossRef]
15. Ezzahri, A.; Boujdi, S.; Bouziani, M.; Yaagoubi, R.; Kenny, L. Soqia-Advice: A Web-GIS Advisory Platform for Efficient Irrigation in Arboriculture. *AgriEngineering* **2024**, *6*, 1594–1618. [CrossRef]
16. Tian, C.; Chen, X.; Chen, J.; Cao, R.; Liu, S. FieldSeg-DA2.0: Further enhancing the spatiotemporal transferability of an individual arable field (IAF) extraction network using multisource remote sensing and land cover data. *Comput. Electron. Agric.* **2024**, *222*, 109050. [CrossRef]

17. Xie, D.; Xu, H.; Xiong, X.; Liu, M.; Hu, H.; Xiong, M.; Liu, L. Cropland Extraction in Southern China from Very High-Resolution Images Based on Deep Learning. *Remote Sens*. **2023**, *15*, 2231. [CrossRef]
18. He, J.; Zeng, W.; Ao, C.; Xing, W.; Gaiser, T.; Srivastava, A.K. Cross-Regional Crop Classification Based on Sentinel-2. *Agronomy* **2024**, *14*, 1084. [CrossRef]
19. Wang, Y.; Zhang, Z.; Zuo, L.; Wang, X.; Zhao, X.; Sun, F. Mapping Crop Distribution Patterns and Changes in China from 2000 to 2015 by Fusing Remote-Sensing, Statistics, and Knowledge-Based Crop Phenology. *Remote Sens*. **2022**, *14*, 1800. [CrossRef]
20. Haldar, D.; Tripathy, R.; Dave, V.; Dave, R.; Bhattacharya, B.K.; Misra, A. Monitoring cotton crop condition through synergy of optical and radar remote sensing. *Geocarto Int*. **2022**, *37*, 377–395. [CrossRef]
21. Wurster, P.M.; Maneta, M.; Kimball, J.S.; Endsley, K.A.; Beguería, S. Monitoring Crop Status in the Continental United States Using the SMAP Level-4 Carbon Product. *Front. Big Data* **2021**, *3*, 597720. [CrossRef]
22. Zhang, Z.; Wang, X.; Zhao, X.; Liu, B.; Yi, L.; Zuo, L.; Wen, Q.; Liu, F.; Xu, J.; Hu, S. A 2010 update of national land use/cover database of China at 1:100000 scale using medium spatial resolution satellite images. *Remote Sens. Environ*. **2014**, *149*, 142–154. [CrossRef]
23. Tiwari, P.S.; Pande, H.; Kumar, M.; Dadhwal, V.K. Potential of IRS P-6 LISS IV for agriculture field boundary delineation. *J. Appl. Remote Sens*. **2009**, *3*, 033528. [CrossRef]
24. Peña-Barragán, J.M.; Ngugi, M.K.; Plant, R.E.; Six, J. Object-based crop identification using multiple vegetation indices, textural features and crop phenology. *Remote Sens. Environ*. **2011**, *115*, 1301–1316. [CrossRef]
25. Zheng, B.; Myint, S.W.; Thenkabail, P.S.; Aggarwal, R.M. A support vector machine to identify irrigated crop types using time-series Landsat NDVI data. *Int. J. Appl. Earth Obs. Geoinf*. **2015**, *34*, 103–112. [CrossRef]
26. Debats, S.R.; Luo, D.; Estes, L.D.; Fuchs, T.J.; Caylor, K.K. A generalized computer vision approach to mapping crop fields in heterogeneous agricultural landscapes. *Remote Sens. Environ*. **2016**, *179*, 210–221. [CrossRef]
27. Tiwari, P.; Poudel, K.P.; Yang, J.; Silva, B.; Yang, Y.; McConnell, M. Marginal agricultural land identification in the Lower Mississippi Alluvial Valley based on remote sensing and machine learning model. *Int. J. Appl. Earth Obs. Geoinf*. **2023**, *125*, 103568. [CrossRef]
28. Bouguettaya, A.; Zarzour, H.; Kechida, A.; Taberkit, A.M. Deep learning techniques to classify agricultural crops through UAV imagery: A review. *Neural Comput. Appl*. **2022**, *34*, 9511–9536. [CrossRef]
29. Zhang, X.; Zhou, Y.; Luo, J. Deep learning for processing and analysis of remote sensing big data: A technical review. *Big Earth Data* **2022**, *6*, 527–560. [CrossRef]
30. Jiao, S.; Hu, D.; Shen, Z.; Wang, H.; Dong, W.; Guo, Y.; Li, S.; Lei, Y.; Kou, W.; Wang, J.; et al. Parcel-Level Mapping of Horticultural Crop Orchards in Complex Mountain Areas Using VHR and Time-Series Images. *Remote Sens*. **2022**, *14*, 2015. [CrossRef]
31. Xu, L.; Ming, D.; Du, T.; Chen, Y.; Dong, D.; Zhou, C. Delineation of cultivated land parcels based on deep convolutional networks and geographical thematic scene division of remotely sensed images. *Comput. Electron. Agric*. **2022**, *192*, 106611. [CrossRef]
32. Marvaniya, S.; Devi, U.; Hazra, J.; Mujumdar, S.; Gupta, N. Small, sparse, but substantial: Techniques for segmenting small agricultural fields using sparse ground data. *Int. J. Remote Sens*. **2021**, *42*, 1512–1534. [CrossRef]
33. Xie, S.; Tu, Z. Holistically-Nested Edge Detection. In *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*; IEEE: Piscataway, NJ, USA, 2015; pp. 1395–1403. [CrossRef]
34. Xia, L.; Liu, R.; Su, Y.; Mi, S.; Yang, D.; Chen, J.; Shen, Z. Crop field extraction from high resolution remote sensing images based on semantic edges and spatial structure map. *Geocarto Int*. **2024**, *39*, 2302176. [CrossRef]
35. Liu, S.; Liu, L.; Xu, F.; Chen, J.; Yuan, Y.; Chen, X. A deep learning method for individual arable field (IAF) extraction with cross-domain adversarial capability. *Comput. Electron. Agric*. **2022**, *203*, 107473. [CrossRef]
36. Xu, L.; Yang, P.; Yu, J.; Peng, F.; Xu, J.; Song, S.; Wu, Y. Extraction of cropland field parcels with high resolution remote sensing using multi-task learning. *Eur. J. Remote Sens*. **2023**, *56*, 2181874. [CrossRef]
37. Wang, S.; Zhou, Y.; Yang, X.; Li, F.; Wu, T.; Luo, J. BSNet: Boundary-semantic-fusion network for farmland parcel mapping in high-resolution satellite images. *Comput. Electron. Agric*. **2023**, *206*, 107683. [CrossRef]
38. Long, J.; Li, M.; Wang, X.; Stein, A. Delineation of agricultural fields using multi-task BsiNet from high-resolution satellite images. *Int. J. Appl. Earth Obs. Geoinf*. **2022**, *112*, 102871. [CrossRef]
39. Diakogiannis, F.I.; Waldner, F.; Caccetta, P.; Wu, C. ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. *ISPRS J. Photogramm. Remote Sens*. **2020**, *162*, 94–114. [CrossRef]
40. Amin, G.; Oberlin, T.; Demarez, V. Early-season delineation of agricultural fields using a fully convolutional multi-task network and satellite images. *Sci. Remote Sens*. **2025**, *12*, 100256. [CrossRef]
41. Li, M.; Long, J.; Stein, A.; Wang, X. Using a semantic edge-aware multi-task neural network to delineate agricultural parcels from remote sensing images. *ISPRS J. Photogramm. Remote Sens*. **2023**, *200*, 24–40. [CrossRef]

42. Long, J.; Zhao, H.; Li, M.; Wang, X.; Lu, C. Integrating Segment Anything Model derived boundary prior and high-level semantics for cropland extraction from high-resolution remote sensing images. *IEEE Geosci. Remote Sens. Lett.* **2024**, *21*, 2505305. [CrossRef]
43. Cai, Z.; Hu, Q.; Zhang, X.; Yang, J.; Wei, H.; Wang, J.; Zeng, Y.; Yin, G.; Li, W.; You, L.; et al. Improving agricultural field parcel delineation with a dual branch spatiotemporal fusion network by integrating multimodal satellite data. *ISPRS J. Photogramm. Remote Sens.* **2023**, *205*, 34–49. [CrossRef]
44. Zhu, M.; Yao, M.; He, Y.; He, Y.; Wu, B. Studies on High-Resolution Remote Sensing Sugarcane Field Extraction based on Deep Learning. *IOP Conf. Ser. Earth Environ. Sci.* **2019**, *237*, 032046. [CrossRef]
45. Yang, R.; Ahmed, Z.U.; Schulthess, U.C.; Kamal, M.; Rai, R. Detecting functional field units from satellite images in smallholder farming systems using a deep learning based computer vision approach: A case study from Bangladesh. *Remote Sens. Appl. Soc. Environ.* **2020**, *20*, 100413. [CrossRef]
46. Crommelinck, S.; Koeva, M.; Yang, M.Y.; Vosselman, G. Application of Deep Learning for Delineation of Visible Cadastral Boundaries from Remote Sensing Imagery. *Remote Sens.* **2019**, *11*, 2505. [CrossRef]
47. Aung, H.L.; Uzkent, B.; Burke, M.; Lobell, D.; Ermon, S. Farm Parcel Delineation Using Spatio-Temporal Convolutional Networks. In *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*; IEEE: Piscataway, NJ, USA, 2020; pp. 340–349. [CrossRef]
48. Qi, L.; Zuo, D.; Wang, Y.; Tao, Y.; Tang, R.; Shi, J.; Gong, J.; Li, B. Convolutional neural network-based method for agriculture plot segmentation in remote sensing images. *Remote Sens.* **2024**, *16*, 346. [CrossRef]
49. Masoud, K.M.; Persello, C.; Tolpekin, V.A. Delineation of Agricultural Field Boundaries from Sentinel-2 Images Using a Novel Super-Resolution Contour Detector Based on Fully Convolutional Networks. *Remote Sens.* **2020**, *12*, 59. [CrossRef]
50. Sun, W.; Sheng, W.; Zhou, R.; Zhu, Y.; Chen, A.; Zhao, S.; Zhang, Q. Deep edge enhancement-based semantic segmentation network for farmland segmentation with satellite imagery. *Comput. Electron. Agric.* **2022**, *202*, 107273. [CrossRef]
51. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. In *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*; IEEE: Piscataway, NJ, USA, 2017; pp. 2980–2988. [CrossRef]
52. Wang, X.; Kong, T.; Shen, C.; Jiang, Y.; Li, L. SOLO: Segmenting Objects by Locations. In *Proceedings of the 16th European Conference on Computer Vision (ECCV)*; Springer: Cham, Switzerland, 2020; pp. 649–665. [CrossRef]
53. Peng, S.; Jiang, W.; Pi, H.; Li, X.; Bao, H.; Zhou, X. Deep Snake for Real-Time Instance Segmentation. In *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: Piscataway, NJ, USA, 2020; pp. 8530–8539. [CrossRef]
54. Zhang, T.; Wei, S.; Ji, S. E2EC: An End-to-End Contour-based Method for High-Quality High-Speed Instance Segmentation. In *Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: Piscataway, NJ, USA, 2022; pp. 4433–4442. [CrossRef]
55. Pan, Y.; Wang, X.; Zhang, L.; Zhong, Y. E2EVAP: End-to-end vectorization of smallholder agricultural parcel boundaries from high-resolution remote sensing imagery. *ISPRS J. Photogramm. Remote Sens.* **2023**, *203*, 246–264. [CrossRef]
56. Zhong, B.; Wei, T.; Luo, X.; Du, B.; Hu, L.; Ao, K.; Yang, A.; Wu, J. Multi-Swin Mask Transformer for instance segmentation of agricultural field extraction. *Remote Sens.* **2023**, *15*, 549. [CrossRef]
57. d'Andrimont, R.; Claverie, M.; Kempeneers, P.; Muraro, D.; Yordanov, M.; Peressutti, D.; Batič, M.; Waldner, F. AI4Boundaries: An open AI-ready dataset to map field boundaries with Sentinel-2 and aerial photography. *Earth Syst. Sci. Data* **2023**, *15*, 317–329. [CrossRef]
58. Wang, S.; Waldner, F.; Lobell, D.B. Unlocking Large-Scale Crop Field Delineation in Smallholder Farming Systems with Transfer Learning and Weak Supervision. *Remote Sens.* **2022**, *14*, 5738. [CrossRef]
59. Chiu, M.T.; Xu, X.; Wei, Y.; Huang, Z.; Schwing, A.G.; Brunner, R.; Khachatrian, H.; Karapetyan, H.; Dozier, I.; Rose, G.; et al. Agriculture-Vision: A Large Aerial Image Database for Agricultural Pattern Analysis. In *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: Piscataway, NJ, USA, 2020; pp. 2825–2835. [CrossRef]
60. Sainte Fare Garnot, V.; Landrieu, L.; Chehata, N. Multi-modal temporal attention models for crop mapping from satellite time series. *ISPRS J. Photogramm. Remote Sens.* **2022**, *187*, 294–305. [CrossRef]
61. Wang, X.; Zhang, R.; Kong, T.; Li, L.; Shen, C. SOLOv2: Dynamic and Fast Instance Segmentation. In *Advances in Neural Information Processing Systems (NeurIPS 2020)*; Curran Associates, Inc.: Red Hook, NY, USA, 2020; Volume 33, pp. 17721–17732.
62. Jocher, G.; Chaurasia, A.; Qiu, J. Ultralytics YOLOv8. 2023. Available online: https://github.com/ultralytics/ultralytics (accessed on 5 September 2026).
63. Ayala, C.; Aranda, C.; Galar, M. Guidelines to compare semantic segmentation maps at different resolutions. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 4702416. [CrossRef]
64. Wei, R.; Yang, L.; Li, X.; Zhu, C.; Zhang, L.; Wang, J.; Liu, J.; Zhu, L.; Zhou, C. Breaking the limitations of scenes and sensors variability: A novel unsupervised domain adaptive instance segmentation framework for agricultural field extraction. *Remote Sens. Environ.* **2025**, *331*, 115051. [CrossRef]

65. Zhou, Z.; Wang, L.; Chen, Q.; Luo, J.; Zhao, X.; Zhang, S.; Zhang, W.; Liao, J.; Lyu, Z. Abandoned land identification in karst mountain area based on time series SAR characteristics at geo-parcels scale. *J. Mt. Sci*. **2023**, *20*, 792–809. [CrossRef]
66. Borowiec, N.; Marmol, U. Using LiDAR System as a Data Source for Agricultural Land Boundaries. *Remote Sens*. **2022**, *14*, 1048. [CrossRef]

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.