

Article

Large-Scale Individual Plastic Greenhouse Extraction Using Deep Learning and High-Resolution Remote Sensing Imagery

Yuguang Chang ^{1,*}, Xiaoyu Yu ^{2,3}, Baipeng Li ³, Xiangyu Tian ^{3,4} and Zhaoming Wu ³ 

¹ School of Civil Engineering, Henan Polytechnic University, Jiaozuo 454000, China

² School of Resources and Environment, Henan Polytechnic University, Jiaozuo 454000, China; shanhua@home.hpu.edu.cn

³ State Key Laboratory of Remote Sensing and Digital Earth, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, China; libp@aircas.ac.cn (B.L.); tianxiangyu20@mails.ucas.ac.cn (X.T.); wuzhaoming21@mails.ucas.ac.cn (Z.W.)

⁴ Key Laboratory of Ecosystem Network Observation and Modeling, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

* Correspondence: changyg@hpu.edu.cn

Abstract

Addressing the demands of agricultural resource digitization and facility crop monitoring, precise extraction of plastic greenhouses using high-resolution remote sensing imagery demonstrates pivotal significance for implementing refined farmland management. However, the complex spatial topological relationships among densely arranged greenhouses and the spectral confusion of ground objects within agricultural backgrounds limit the effectiveness of conventional methods in the large-scale, precise extraction of plastic greenhouses. This study constructs an Individual Plastic Greenhouse Extraction Network (IPGENet) by integrating a multi-scale feature fusion decoder with the Swin-UNet architecture to improve the accuracy of large-scale individual plastic greenhouse extraction. To ensure sample accuracy while reducing manual labor costs, an iterative sampling approach is proposed to rapidly expand a small sample set into a large-scale dataset. Using GF-2 satellite imagery data in Shandong Province, China, the model realized large-scale mapping of individual plastic greenhouse extraction results. In addition to large-scale sub-meter extraction and mapping, the study conducted quantitative and spatial statistical analyses of extraction results across cities in Shandong Province, revealing regional disparities in plastic greenhouse development and providing a novel technical approach for large-scale plastic greenhouse mapping.

Keywords: agricultural monitoring; individual plastic greenhouses; GF-2 remote sensing imagery; feature fusion; semantic segmentation



Academic Editor: Mercedes Del Río Celestino

Received: 7 July 2025

Revised: 2 August 2025

Accepted: 18 August 2025

Published: 21 August 2025

Citation: Chang, Y.; Yu, X.; Li, B.; Tian, X.; Wu, Z. Large-Scale Individual Plastic Greenhouse Extraction Using Deep Learning and High-Resolution Remote Sensing Imagery. *Agronomy* **2025**, *15*, 2014. <https://doi.org/10.3390/agronomy15082014>

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

1. Introduction

As a vital component of protected agriculture, greenhouses mitigate the impact of natural environmental factors, such as weather and temperature, on crops through intelligent management with climate control technologies, thereby enhancing crop yields and playing a critical role in global precision agriculture development [1]. Protected horticulture using plastic greenhouses initially emerged in China during the 1960s, primarily applied in vegetable cultivation [2]. In 2023, China's greenhouse cultivation area accounted for over 70% of the global total, reaching 18.3587 million hectares, with concentrated cultivation zones formed in provinces such as Shandong and Yunnan [2]. The rational integration of

greenhouse cultivation into vertical farming systems not only enhances crop yield per unit area for vegetables and fruits but also effectively conserves land resources, offering a viable pathway to address the global food security crisis [3–5]. While greenhouse cultivation can partially address food security and economic problems, the environmental issues it brings cannot be ignored [6]. The predominant structure in greenhouse cultivation is plastic greenhouses, top-covered with synthetic polymers resistant to natural degradation [7–9]. Without proper disposal, these plastic films will cause “white pollution” to the environment, resulting in reduced biodiversity [10–12]. Therefore, accurate statistical analysis and mapping of greenhouse quantity and area within the region are not only critical for the development of local horticultural crop production, but also profoundly impact the sustainable development of natural ecosystems.

The rapid advancement of remote sensing technology has enhanced object recognition and other applications with abundant data support [13], thereby providing more methodologies for greenhouse extraction. Early feature extraction methods primarily focused on studying the spectral characteristics of target features, utilizing machine learning approaches for feature extraction [14]. Yuan et al. (2013) integrated spectral and textural features of ground objects in remote sensing imagery with partial spectral histograms for feature combination, proposing a remote sensing image segmentation method that achieved 75.1% accuracy on the IKONOS dataset, though performance degradation occurred in complex scenarios [15]. Sisodia et al. (2014) utilized supervised maximum likelihood classification (MLC) to categorize five land cover types (water bodies, hills, wasteland, urban areas, and vegetation) on Landsat ETM+ imagery in Jaipur town, India, achieving an overall accuracy of 93.75% [16]. Among various feature extraction methods, Luo et al. (2025) employed a pixel-based random forest classifier as the greenhouse classification method, trained using spectral, index, and texture features, to classify greenhouses across six Chinese provinces (Guizhou, Hunan, Jiangxi, Fujian, Guangdong, and Guangxi), achieving an overall accuracy of 96.48% [17]. Chen et al. (2025) conducted spectral analysis of plastic and non-plastic greenhouses in the study areas of Weifang, Nantong, and Kunming (China), as well as Da Lat (Vietnam), based on Landsat-8 Operational Land Imager (OLI) spectral features, achieving a maximum F1-score of 87.9% [18]. However, the classification efficacy for roads and gray-colored greenhouses remained poor. Sun et al. (2025) employed the Google Earth Engine cloud platform and Landsat 7 remote sensing imagery to classify agricultural greenhouses across China using the random forest method, creating a national classification dataset for the years 2010, 2016, and 2022 [19].

Deep learning is a computational technology that mimics human neural networks [20]. The rapid development of deep learning techniques in recent years, combined with remote sensing imagery, has significantly enhanced the precision of feature recognition and classification, supported by massive datasets [21–23]. The advancement of deep learning models has not only minimized manual intervention in remote sensing object recognition and classification but also significantly enhanced accuracy rates [24]. Tong et al. (2024) combined segmentation and classification frameworks using CNN methods to map global greenhouse facilities in 2019, revealing a surge in cultivation and its multifaceted impacts [25]. Tian et al. (2023) addressed challenges including small farmland plots and complex crop compositions by integrating deep learning networks with object-oriented classification methods, achieving an F1-score of 84.49% for crop classification and mapping in Xishui County, Guizhou Province, China, using multi-source remote sensing imagery [26]. Xie et al. (2024) developed the RSWDet network by integrating a dual-branch structured point detection head and a low-level feature enhancement module, achieving an 83.1% mean accuracy rate for large-scale wind turbine detection using GF-2 remote sensing imagery [27]. In the integration of deep learning techniques and remote sensing imagery, numerous

researchers have utilized this approach for greenhouse recognition and extraction. Chen et al. (2020) [28] proposed a dual-task learning module to enhance the base CNN network, utilizing Google Earth imagery for nationwide greenhouse extraction and achieving 2% higher average accuracy than Mask R-CNN [29]. Wang et al. (2023) established an Area and Quantity Extraction Framework (AQSEF) by integrating UNet with YOLO v5, achieving 91.4% accuracy in extracting plastic greenhouse coverage area and quantity in Beijing, China [30]. Liu et al. (2024) proposed improvements to the YOLOX deep learning model for identifying plastic greenhouses in Weifang City, Shandong Province, China, and statistically analyzing their quantity and area [31].

Despite significant advances in greenhouse mapping achieved by numerous scholars using deep learning techniques, current province-scale mapping methodologies fail to accurately delineate structural boundaries and spatial–topological relationships of individual plastic greenhouses. This limitation introduces substantial errors in quantity statistics, morphological distortions, and misclassification risks, severely impeding data-driven agricultural management decisions and digital transformation across the industry supply chain. To address the aforementioned issues, this study proposes a deep learning-based method utilizing high-resolution remote sensing imagery for the precise extraction of monolithic plastic greenhouses at provincial-scale regions. This method integrates sample iteration with multi-scale feature fusion techniques to enhance the extraction accuracy of large-scale individual plastic greenhouses across Shandong Province. The primary contributions of this study are as follows:

- A Multi-scale Feature Fusion Decoder (MFFD) module was designed. By upsampling high-level semantic features and concatenating them with low-level detail features, followed by convolutional processing, it enhances contextual awareness and strengthens the network's capability to delineate edges of individual plastic greenhouses.
- An Individual Plastic Greenhouse Extraction Network (IPGENet) was constructed. The Swin-UNet baseline architecture was improved by designing an MFFD (Multi-scale Feature Fusion and Distillation) module to enhance its capability for individual plastic greenhouse extraction.
- An iterative sample method is proposed. Through sample iteration based on the IPGENet framework, efficient dataset expansion is achieved starting from a limited initial sample set. This method ensures sample labeling accuracy while substantially reducing manual annotation costs and drastically improving the efficiency of large-scale sample dataset construction.
- High-precision mapping of individual plastic greenhouses throughout Shandong Province. The deep learning-based framework for individual plastic greenhouse extraction was systematically validated through experiments. Utilizing GF-2 remote sensing imagery, it achieved sub-meter-level extraction and mapping of plastic greenhouses across Shandong Province, realizing geometric accuracy recognition at the sub-meter scale for individual structures.

The remainder of this paper is structured as follows: Section 2 details the research area, dataset construction, and the architecture of the feature extraction network. Section 3 presents the model performance comparison, large-scale mapping of individual plastic greenhouses across Shandong Province, and statistical analysis. Section 4 discusses the strengths, limitations, and future research directions of this study. Section 5 systematically summarizes the present study.

2. Materials and Methods

2.1. Study Area

The study area covers the entire Shandong Province, China, aiming to achieve high-precision extraction and mapping of large-scale individual plastic greenhouses. Shandong Province is situated in the eastern coastal region of China and the lower reaches of the Yellow River ($34^{\circ}23' \text{ N}$ – $38^{\circ}17' \text{ N}$, $114^{\circ}48' \text{ E}$ – $122^{\circ}42' \text{ E}$), bordering the Yellow Sea and Bohai Sea to the east. The total land area of the province is 15.81 million square kilometers. The topography of Shandong Province is predominantly composed of plains and hills (as shown in Figure 1), with the low-lying terrain of the Yellow River Delta in the northwest. The warm temperate monsoon climate, characterized by distinct seasonal variations, has fostered the development of advanced agricultural systems and modern industrial sectors. In the 1990s, Shandong Province initiated large-scale promotion of plastic greenhouse cultivation, with Shouguang City, which is renowned as the “Vegetable Capital of China”, evolving into a national benchmark for protected agriculture development. Shandong Province exhibits extensive greenhouse coverage, high distinguishability in remote sensing imagery, and strong application demands, which collectively establish a robust foundation for developing and validating high-precision remote sensing extraction models.

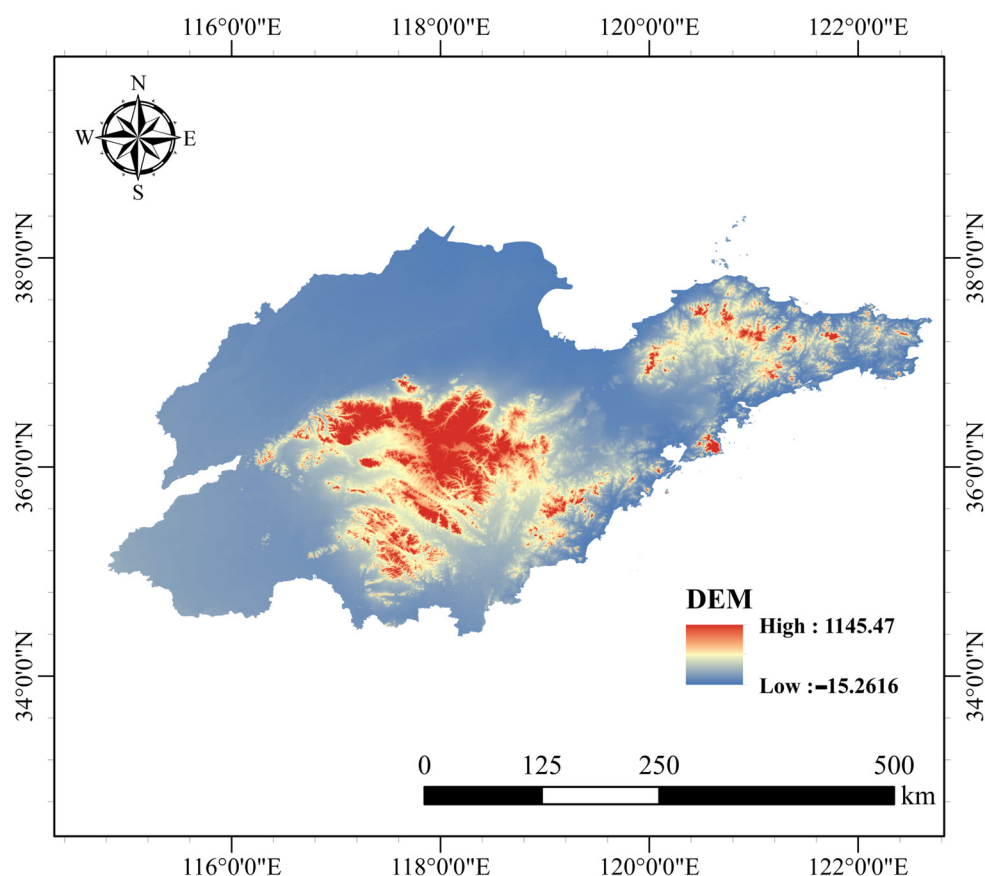


Figure 1. Elevation map of Shandong Province, China.

2.2. Datasets

Shandong Province has transformed its modern agricultural landscape through plastic greenhouse technology, characterized by high-density deployment and large-scale coverage across the region. This study develops a deep learning model using domestic GF-2 satellite imagery (spectral range and resolution detailed in Table 1) to achieve province-wide, sub-meter spatial precision extraction of plastic greenhouses in Shandong, China. The image

acquisition period in this study primarily spanned April and December 2020, covering parts of Shouguang, Hanting, Weicheng, and Changle districts in Weifang City; parts of Juye, Chengwu, and Shanxian counties in Heze City; and parts of Jinxiang County in Jining City, Shandong Province. The total coverage area is 1687.05 km². The GF-2 remote sensing imagery comprises 0.8 m-resolution panchromatic imagery and 3.2 m-resolution multispectral imagery, with their spectral ranges detailed in Table 1. To facilitate manual annotation, this study performed cloud removal and color enhancement on both panchromatic and multispectral imagery during the preprocessing phase, followed by fusion of both imagery.

Table 1. Resolution and spectral range of GF-2 remote sensing imagery.

	Sensor Resolution	Spectral Range
Panchromatic imagery	0.8 m	450–900 nm
Multispectral imagery	3.2 m	450–520 nm (Blue band)
		520–590 nm (Green band)
		630–690 nm (Red band)
		770–890 nm (Near-infrared band)

Figure 2 illustrates representative greenhouse typology cases observed in GF-2 imagery, serving as critical references for the annotation workflow to ensure consistent labeling of individual greenhouses and enhance the robustness of model training and evaluation. In this study, a total of 68,141 manually annotated labels were expanded to 145,582 training samples of 512×512 images with corresponding labels through cropping, rotation, and flipping operations. From these, 15,000 images and corresponding labels were randomly selected for ablation and comparative experiments.

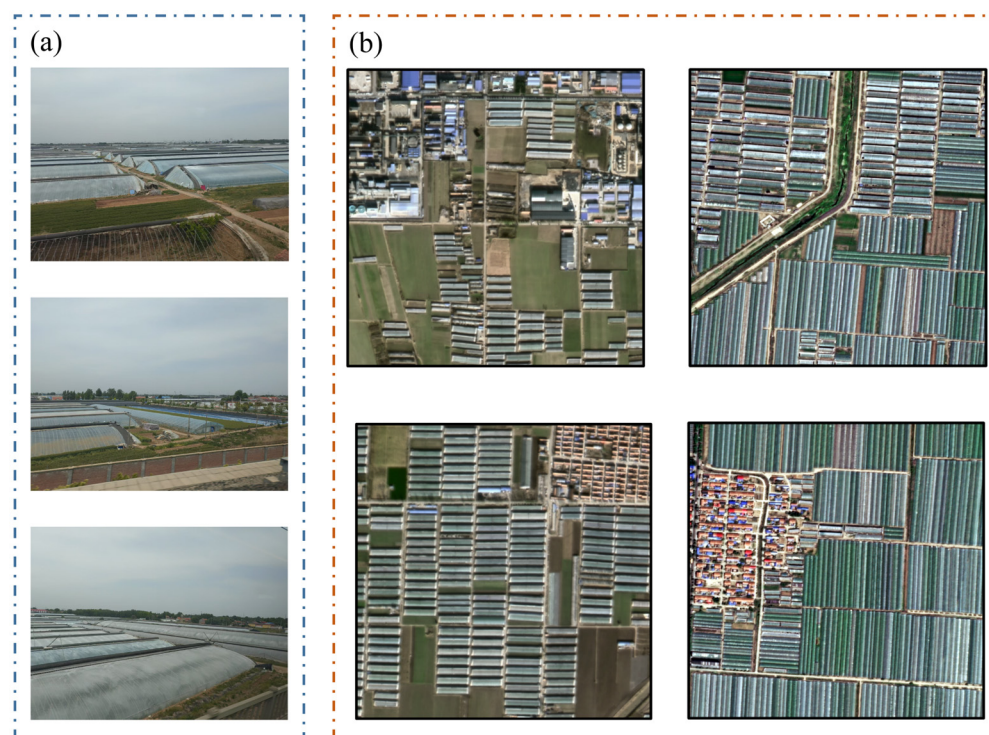


Figure 2. Illustration of greenhouse structural types, where (a) shows the actual appearance of plastic greenhouses, and (b) presents their manifestation on GF-2 remote sensing imagery.

For the dataset covering the entire Shandong Province for individual plastic greenhouse extraction, a total of 145,582 images with corresponding labels were divided into training and

validation sets at an 8:2 ratio. The training set comprised 116,466 images and corresponding labels, while the validation set contained 29,116 images and corresponding labels.

2.3. Methods

The extraction workflow for individual plastic greenhouses throughout Shandong Province is illustrated in Figure 3. In this study, the extraction process of single-span plastic greenhouses was divided into two main components: (1) dataset preparation, and (2) model training and result extraction.

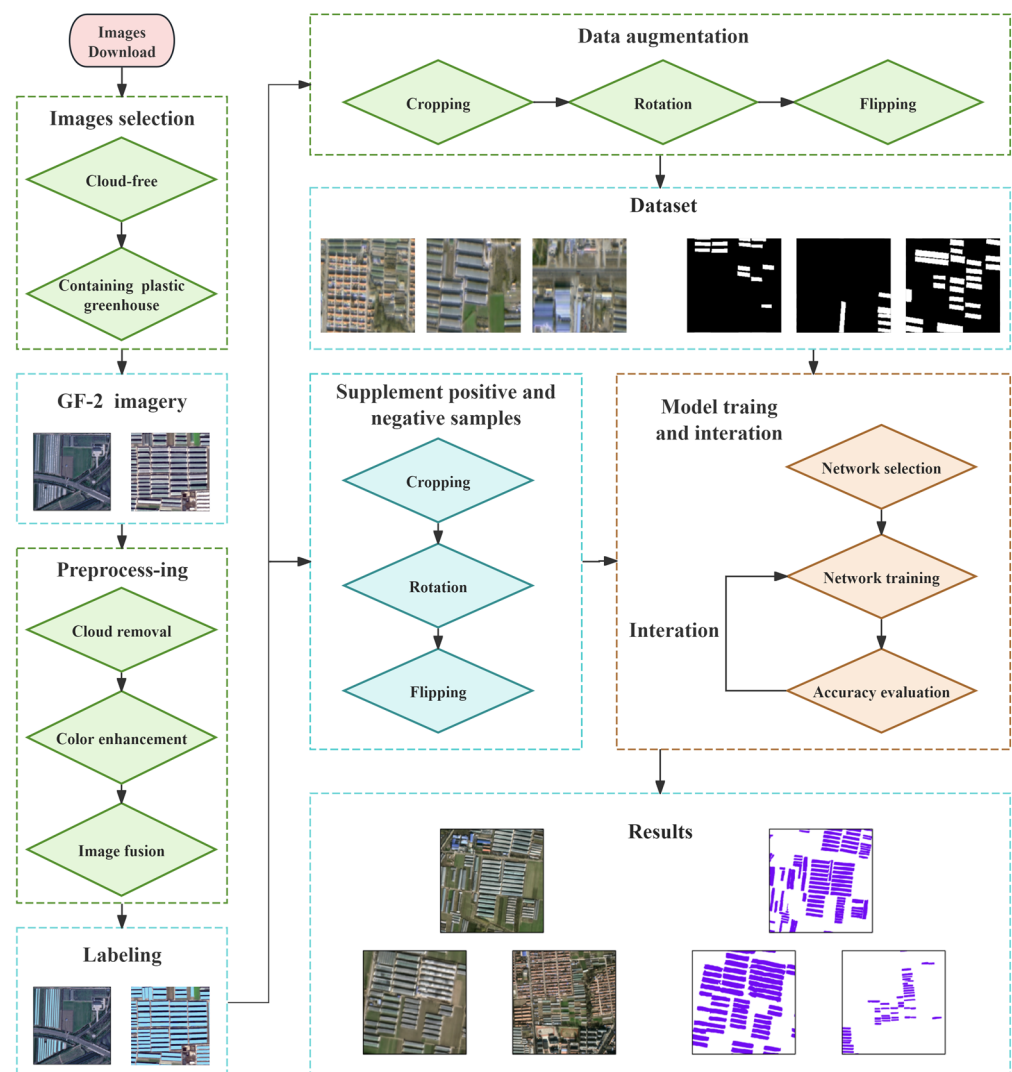


Figure 3. Extraction workflow for individual plastic greenhouses across Shandong Province.

During the dataset preparation phase, this study applies a 10% cloud cover threshold standard to screen Gaofen-2 (GF-2) satellite imagery within Shandong Province. Subsequently, preprocessing operations, including cloud removal, color enhancement, and image fusion, were performed on the imagery. Following preprocessing, vector labels of individual plastic greenhouses were obtained through manual sample annotation on the preprocessed remote sensing images using professional remote sensing image processing software, and subsequently converted into raster format. After obtaining raster-format human-annotated labels, both the labels and corresponding remote sensing images underwent cropping, rotation, and flipping operations to generate 512×512 pixel image patches. These patches were ultimately partitioned into training and validation sets at an 8:2 ratio.

During the model training and result extraction phase, this study employed the proposed IPGENet (Individual Plastic Greenhouse Extraction Network) for model training on both the training and validation sets. Through iterative optimization, optimal model parameters were obtained. The model trained through iterative optimization was then applied to GF-2 imagery of Shandong Province, yielding preliminary extraction results for individual plastic greenhouses. Through comparative analysis between initial extraction results and corresponding remote sensing imagery, instances of mis-extraction and omission were identified. Consequently, additional positive and negative samples were incorporated to retrain and iteratively refine the model, ultimately achieving significantly improved province-wide extraction results for individual plastic greenhouses in Shandong Province.

The experiments were conducted using the PyTorch (2.0.1) framework [32], with four NVIDIA GeForce RTX 3090 GPUs operating under a CUDA 11.6 environment. The training epochs were set to 100 with a batch size of 8, and AdamW [33] was employed as the optimizer.

2.3.1. IPGENet Architecture

The IPGENet architecture used in this study is shown in Figure 4. Swin Transformer exhibits powerful feature representation capabilities. It was initially applied to image segmentation tasks in medical image analysis, achieving outstanding performance [34,35], and holds significant importance in image-guided clinical procedures. The IPGENet used in this study is based on the Swin-UNet framework [34], employing a Swin Transformer as the backbone for feature extraction and utilizing the proposed Multi-scale Feature Fusion Decoder (MFFD) module as the decoder. During the feature extraction stage of IPGENet, the feature layers extracted by the Swin Transformer undergo separation of the deepest layer features from the remaining layers through the “Deepest Layer Feature Removal” mechanism. Spatial Pyramid Pooling (SPP) is applied to the deepest features to enhance global contextual comprehension; for the remaining feature layers, a “Feature Layer Order Reversal” strategy reverses their sequence from shallow-to-deep to deep-to-shallow. The structural design of IPGENet meets the hierarchical feature requirements for the subsequent Multi-scale Feature Fusion Decoder (MFFD) module, enhancing compatibility between deep and shallow features while mitigating information redundancy.

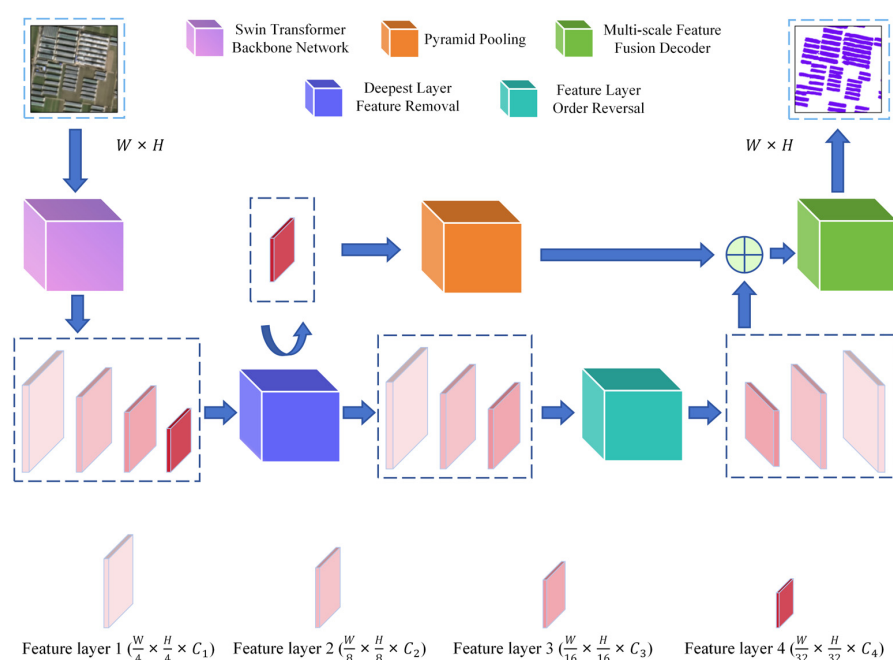


Figure 4. Individual Plastic Greenhouse Extraction Network (IPGENet) architecture.

2.3.2. Multi-Scale Feature Fusion Decoder (MFFD)

Prior researchers have improved networks using Multi-scale Feature Fusion as a research direction. Based on residual learning, Qin et al. (2020) designed a Multi-Scale Feature Fusion Residual Block (MSFFRB) capable of adaptively detecting and fusing multi-scale image features [36]; Zhang et al. (2020) proposed a Weighted Feature Kernels Convolutional Neural Network (WFCNN), which employs an encoder to extract multi-level spectral and semantic features. Through linear fusion and refinement layers, it optimizes feature stability, while hierarchically integrating multi-scale semantic and spectral features to construct high-robustness feature maps [37]; Wang et al. (2023) employed a channel attention mechanism to capture feature maps with varying receptive fields, accomplishing cross-channel and spatial dimension feature fusion through an integrated pyramid module [38]. Inspired by extensive research on Multi-scale Feature Fusion, we propose a Multi-scale Feature Fusion Decoder (MFFD) module, which enhances multi-scale feature representation and detail preservation through cross-layer feature concatenation, window attention mechanism optimization, and dynamic resolution adaptation. The modular structure of the MFFD proposed in this study is illustrated in Figure 5.

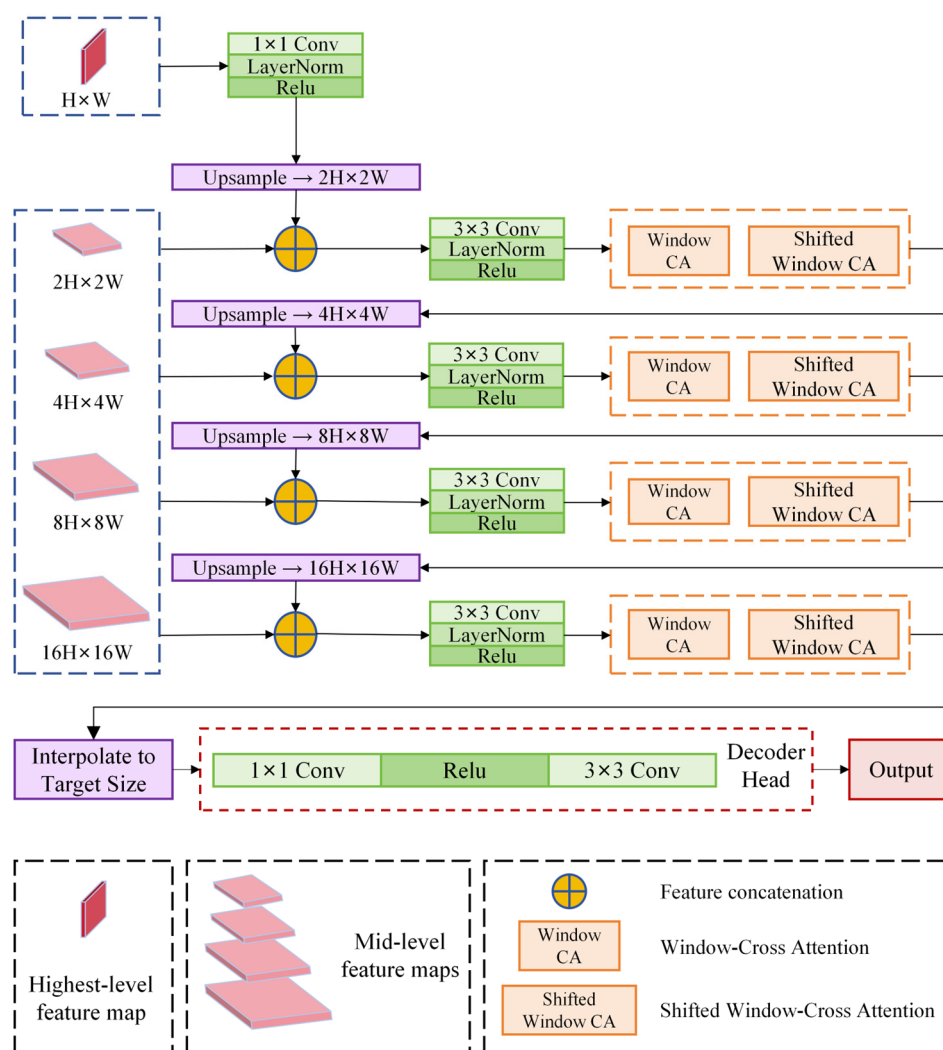


Figure 5. MFFD module architecture diagram.

The primary workflow of the MFFD module designed in this study is as follows:

- (1) Initial processing:

$$M_0 = \text{ReLU}(\text{LayerNorm}(\text{Conv}_{1 \times 1}(M_{\text{init}}))). \quad (1)$$

First, a 1×1 lateral convolution is applied to the highest-level feature map. In the formula, M_{init} denotes the initial deepest feature layer to be processed, $\text{Conv}_{1 \times 1}$ represents the 1×1 lateral convolution, LayerNorm indicates normalization processing, ReLU signifies the ReLU activation function, and M_0 denotes the features processed at the initial stage.

(2) Multi-scale Feature Fusion ($i = 0, 1, 2, 3$):

$$M_i^u = \text{Upsample}_{\text{bilinear}}(M_i, x_i), \quad (2)$$

$$f_i = \text{ReLU}(\text{LayerNorm}(\text{Conv}_{3 \times 3}(\text{Concat}(x_i, M_i^u)))), \quad (3)$$

$$M_{\text{flat}} = \text{Flatten}(M_i^u) \in \mathbb{R}^{B \times (H_i \cdot W_i) \times D}, \quad (4)$$

$$f_{i,\text{flat}} = \text{Flatten}(f_i) \in \mathbb{R}^{B \times (H_i \cdot W_i) \times D}, \quad (5)$$

$$M_{\text{attn}} = \text{SWCA}_{\text{seq}}^{(i)}(M_{\text{flat}}, f_{i,\text{flat}}, (H_i, W_i)), \quad (6)$$

$$M_{i+1} = \text{Reshape}(M_{\text{attn}}, (B, D, H_i, W_i)). \quad (7)$$

The multi-scale feature fusion phase consists of four consecutive processing steps, with sequential feature handling procedures as follows:

Stage 1: First apply 1×1 lateral convolution to the highest-level feature map, then perform bilinear upsampling to $2H \times 2W$ resolution, followed by feature concatenation with the mid-level feature maps at $2H \times 2W$ resolution. Following feature concatenation, a 3×3 lateral convolution is applied to the concatenated features, followed by alternating utilization of “Window CA” and “Shifted Window CA” to overcome fixed-window limitations and reduce computational complexity. Stage 2: The results from Stage 1 are upsampled to dimensions $4H \times 4W$ via bilinear interpolation and fused by feature concatenation with $4H \times 4W$ mid-level feature maps. The concatenated features undergo 3×3 lateral convolution, followed by alternately applied Window CA and Shifted Window CA. Stage 3: Following a process analogous to Stage 2, the results from Stage 2 are upsampled to $8H \times 8W$ resolution via bilinear interpolation and subsequently concatenated with $8H \times 8W$ mid-level feature maps. The concatenated features are then processed through a 3×3 lateral convolution, followed by alternating applications of “Window CA” (Window Cross-Attention) and “Shifted Window CA” (Shifted Window Cross-Attention). Stage 4: Following a similar procedure to Stage 3, the results from Stage 3 are upsampled to $16H \times 16W$ resolution via bilinear interpolation and concatenated with $16H \times 16W$ mid-level feature maps through feature concatenation. The concatenated features undergo 3×3 lateral convolution, followed by alternating application of “Window CA” and “Shifted Window CA”. In the formula, x_i denotes the intermediate feature layer at the current scale input, Concat represents feature concatenation processing, $\text{Conv}_{3 \times 3}$ indicates 3×3 lateral convolution, Flatten signifies sequence transformation, $\text{SWCA}_{\text{seq}}^{(i)}$ corresponds to cross-attention processing, comprising both Window-Cross Attention and Shifted Window-Cross Attention modules, and M_{i+1} denotes feature restoration with spatial dimension reconstruction.

(3) Final Output Processing:

$$M_{out} = Upsample_{bilinear'} \quad (8)$$

$$y = Conv_{3 \times 3}(ReLU(Conv_{1 \times 1}(M_{out}))). \quad (9)$$

Finally, after four consecutive feature processing stages, the target features are interpolated to the target dimensions and passed through the decoder head for 1×1 convolution, followed by 3×3 convolution to generate the output results. In the formula, M_{out} denotes the upsampled feature map to target dimensions, and y represents the final output results.

The MFFD module concatenates the upsampled deepest feature layer with the corresponding intermediate feature layer at each stage, followed by a 3×3 lateral convolution operation, which strengthens feature fusion capability.

2.3.3. Accuracy Evaluation Metrics

This study will measure model efficacy through four core dimensions: Recall, Precision, Intersection over Union(IoU) and the comprehensive balanced metric (F1-score). This study systematically evaluates the comprehensive performance of models in single plastic greenhouse extraction tasks through four core metrics (Recall, Precision, IoU and F1-score), quantitatively analyzes their precision advantages and scenario adaptability limitations, and drives targeted optimization of model architectures to enhance the accuracy and regional adaptability of large-scale greenhouse extraction tasks in Shandong Province. Recall and precision serve as the dual core performance metrics for binary classification models, characterizing the model's coverage capability of true positive instances and the reliability of prediction results, respectively. The Intersection over Union (IoU) metric quantifies spatial localization consistency by computing the ratio of overlapping area between the predicted bounding box and ground truth annotation (intersection divided by union). A value approaching 1 indicates excellent spatial alignment between the predicted region and actual distribution. The F1-score establishes a dynamic equilibrium metric for model efficacy in imbalanced classification scenarios through its harmonic mean mechanism between precision and recall. Its mathematical essence resides in the nonlinear weighting scheme that simultaneously constrains the risks of false positives (FP) and false negatives (FN).

$$\text{Recall} = \frac{TP}{TP + FN'} \quad (10)$$

$$\text{Precision} = \frac{TP}{TP + FP'} \quad (11)$$

$$\text{IoU} = \frac{TP}{TP + FP + FN'} \quad (12)$$

$$\text{F1-score} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}. \quad (13)$$

TP (True Positive) represents cases where the remote sensing image contains a greenhouse and the model correctly predicts it as a greenhouse. TN (True Negative) denotes instances where the remote sensing image shows non-greenhouse land covers and the model accurately identifies them as other categories. FP (False Positive) refers to scenarios where the image contains non-greenhouse features but the model erroneously classifies them as greenhouses. FN (False Negative) indicates situations where a greenhouse exists in the image, but the model fails to detect it and misclassifies it as other land cover types.

3. Results

3.1. Ablation Study

3.1.1. Quantitative Comparisons

To validate the performance improvement of the proposed IPGENet architecture and MFFD module over Swin-UNet, comparative experiments were conducted between IPGENet and Swin-UNet on the self-constructed dataset. In the comparative experiments, Swin-UNet was employed as the baseline network, followed by incorporating the MFFD (Multi-scale Feature Fusion Decoder) module to conduct performance comparisons. The specific experimental metrics are detailed in Table 2.

Table 2. Comparative evaluation metrics between Swin-UNet and IPGENet.

Methods	Recall (%)	Precision (%)	IoU (%)	F1-Score (%)
Baseline	91.24	91.13	83.82	91.19
+MFFD	93.46	92.86	87.45	93.16

IPGENet achieved 93.46% recall and 92.86% precision on the dataset, surpassing Swin-UNet's performance of 91.24% (recall) and 91.13% (precision). This demonstrates IPGENet's superior reliability in greenhouse coverage identification and prediction on remote sensing imagery, with fewer non-greenhouse areas misclassified as greenhouse regions compared to Swin-UNet. The IoU (Intersection over Union) metric of IPGENet was significantly higher than that of Swin-UNet, indicating stronger spatial consistency between predicted greenhouse areas and ground truth annotations in IPGENet compared to Swin-UNet. The F1-score reflects the comprehensive classification performance of the model on target features. IPGENet achieved a 93.16% F1-score on our custom dataset, surpassing Swin-UNet and demonstrating its exceptional classification capability. Comparative experimental results validated the effectiveness of the MFFD module in enhancing Swin-UNet performance, demonstrating IPGENet's improved capability for identifying and extracting large-scale individual plastic greenhouses in Shandong Province.

3.1.2. Extraction Result Comparison

To visually demonstrate the performance comparison between IPGENet and Swin-UNet, representative extraction results of individual plastic greenhouses obtained by both methods are presented in Figure 6. This study compares three categories of individual plastic greenhouse extraction results: (1) those coexisting with farmland, (2) those coexisting with buildings, and (3) densely distributed clusters. Each group contains two types of individual plastic greenhouses. One type is characterized by long and narrow single-span plastic greenhouses; the other features short and wide structures with larger spacing (red indicates erroneous extraction, while blue denotes omission extraction.)

In the extraction results of Group (1), IPGNet and Swin-UNet demonstrated comparable performance. However, Swin-UNet yielded several false positives where roads were misidentified as individual plastic greenhouses.

In the extraction results of Group (2), Swin-UNet exhibited omission errors for elongated and narrow individual plastic greenhouses, while erroneous adhesion occurred in the extracted results for short and wide greenhouse types. In contrast, IPGENet achieved superior performance in extracting both types of individual plastic greenhouses.

In the extraction results of Group (3) for individual plastic greenhouses, densely distributed long and narrow greenhouses were observed. Under such conditions, both Swin-UNet and IPGENet exhibited omission errors; however, mis-extractions were additionally present in the Swin-UNet results. For short and wide, large-spaced individual

plastic greenhouses, IPGENet extraction results exhibited only minor false positives, while the Swin-UNet results showed significantly more false positives alongside minor omissions.

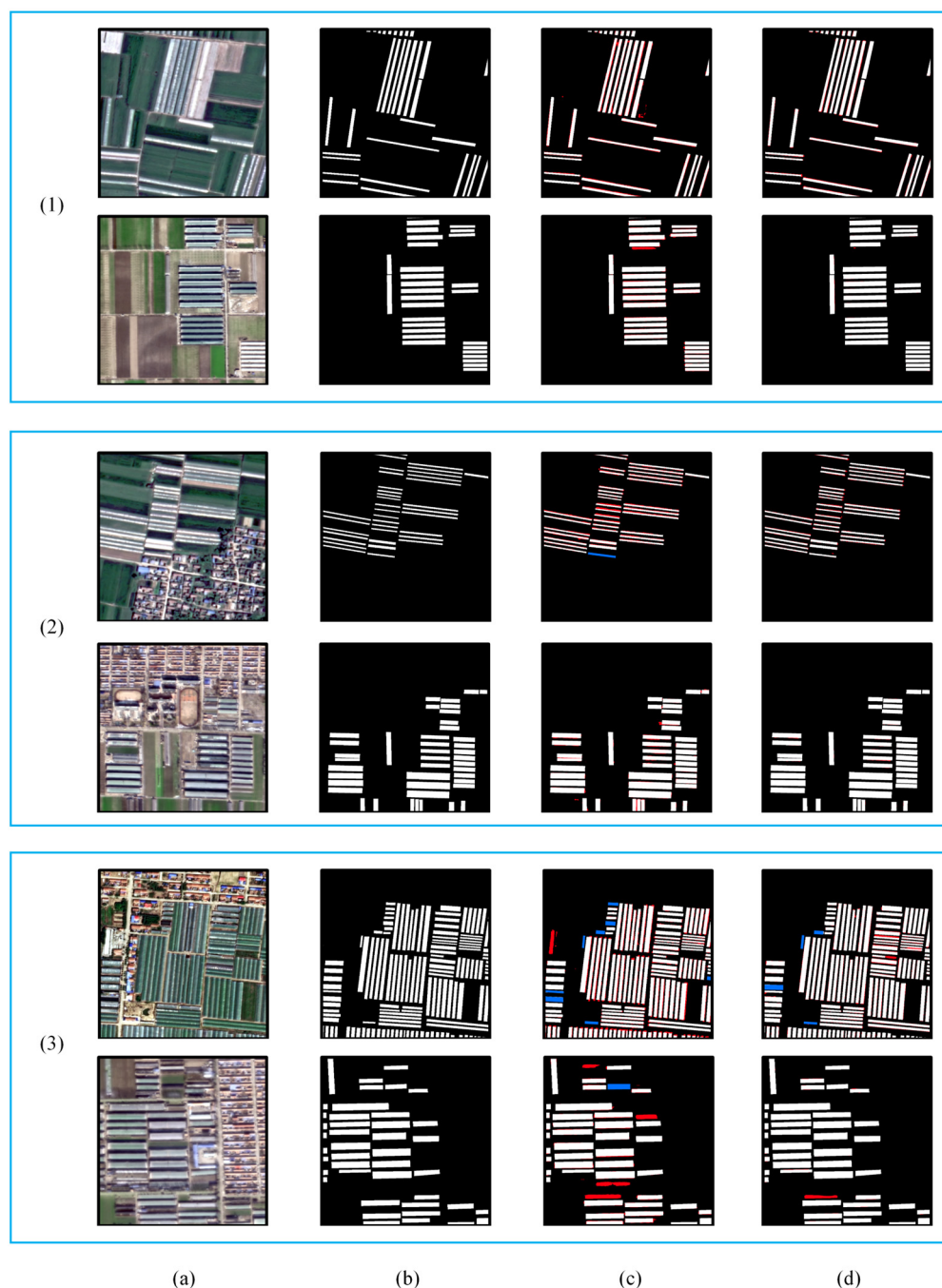


Figure 6. Comparison of extraction results between Swin-UNet and IPGENet. (a) GF-2 remote sensing image. (b) Label corresponding to the remote sensing image. (c) Extraction results of individual plastic greenhouses using Swin-UNet. (d) Extraction results of individual plastic greenhouses using IPGENet. (1) Group 1: Individual plastic greenhouses coexisting with farmland. (2) Group 2: Individual plastic greenhouses coexisting with buildings. (3) Group 3: Densely distributed individual plastic greenhouses.

Comprehensive quantitative analysis and comparative extraction results demonstrate that the proposed IPGENet outperforms the Swin-UNet network, with the MFFD module significantly enhancing Swin-UNet’s ability to reduce false and missed extractions.

3.2. Comparison of Methods

3.2.1. Quantitative Comparisons

In addition to comparative trials between IPGENet and Swin-UNet, this study also contrasts IPGENet with other widely adopted deep learning networks (experimental results are shown in Table 3), such as UNet [39], UperNet [40], PAN [41], and Res-UNet [42]. UNet adopts a U-shaped symmetric encoder–decoder architecture, where skip connections fuse multi-scale features to precisely localize target contours. With its compact parameterization, it demonstrates strong adaptability for small-sample training. UPerNet enhances multi-object recognition by constructing a feature pyramid that integrates low-level spatial details with high-level semantics, while dynamically sampling heterogeneous annotation types during training. PAN (Path Aggregation Network) enhances traditional feature pyramid networks by introducing a bottom-up information pathway that complements the top-down structure of standard FPNs. This bidirectional feature fusion architecture improves multi-scale feature synergy, thereby enhancing localization precision and semantic representation for detection and segmentation tasks. Res-UNet integrates residual connections into the U-Net architecture, employing residual modules in the encoder to facilitate feature propagation and mitigate gradient vanishing in deeper networks. Its decoder utilizes skip connections to fuse spatial detail with semantic context, demonstrating superior performance in low-contrast scenes and small-object segmentation.

Table 3. Test accuracy table of widely used semantic segmentation methods. Bold values represent the maximum values, and underlined values indicate the second-highest values.

Methods	Recall (%)	Precision (%)	IoU (%)	F1-Score (%)
UNet	91.12	91.07	82.80	91.09
UperNet	91.09	90.46	82.19	90.77
PAN	92.31	91.89	85.72	92.31
Res-UNet	<u>92.98</u>	<u>92.77</u>	<u>86.69</u>	<u>92.87</u>
Swin-UNet	91.24	91.13	83.82	91.19
Our	93.46	92.86	87.45	93.16

The experimental results of commonly used deep learning networks on the custom dataset are summarized in Table 3. Res-UNet achieved an F1-score of 92.87% among the conventional methods, demonstrating the best overall performance. Although UperNet exhibits comparable recall rates to UNet and Swin-UNet, its precision and IoU values are the lowest at 90.46% and 82.19%, respectively. This indicates UperNet demonstrates the weakest coverage capability for individual plastic greenhouses in remote sensing imagery, the lowest reliability in prediction results, and the poorest spatial alignment between detected locations and actual distributions. Compared to commonly used deep learning methods, the proposed IPGENet demonstrates superior performance across all evaluation metrics, achieving a comprehensive F1-score of 93.16%, recall of 93.46%, precision of 92.86%, and Intersection over Union (IoU) of 87.45%.

3.2.2. Extraction Result Comparison

In addition to conducting a quantitative comparison and analysis between IPGENet and other commonly used deep learning networks, this study also demonstrates the extraction results of single plastic greenhouses for each deep learning approach. The specific demonstration results are shown in Figure 7. This study compares three categories of individual plastic greenhouse extraction results: (1) those coexisting with farmland, (2) those coexisting with buildings, and (3) densely distributed clusters. Each group contains two types of individual plastic greenhouses. One type is characterized by long and

narrow single-span plastic greenhouses; the other features short and wide structures with larger spacing (red indicates erroneous extraction, while blue denotes omission extraction.)



Figure 7. Comparison of extraction results between common deep learning networks and IPGENet for individual plastic greenhouses. (a) GF-2 remote sensing image. (b) Label corresponding to the remote sensing image. (c) UNet. (d) UPerNet. (e) PAN. (f) Res-UNet. (g) Swin-UNet. (h) IPGENet. (1) Group 1: Individual plastic greenhouses coexisting with farmland. (2) Group 2: Individual plastic greenhouses coexisting with buildings. (3) Group 3: Densely distributed individual plastic greenhouses.

For elongated narrow individual plastic greenhouses in the Group (1) extraction results, erroneous extractions were observed with UNet, significant adhesion occurred with UPerNet, and minor erroneous extractions with adhesion appeared in the Swin-UNet results, while other networks exhibited similar performance. For short and wide individual plastic greenhouses, Unet, UperNet, PAN, and Swin-UNet were observed to misclassify certain road sections as plastic greenhouses, while Res-Unet and IPGENet demonstrated superior extraction performance.

For elongated narrow individual plastic greenhouses in the Group (2) extraction results, the following performance variations were observed across different models: UNet exhibited omission errors and minor false extractions for elongated and narrow individual plastic greenhouses; UPerNet demonstrated significant adhesion phenomena among adjacent instances; PAN and Swin-UNet primarily showed omission errors, while Res-UNet and IPGENet achieved optimal extraction performance. For short and wide-span

individual plastic greenhouses, the extraction results achieved the desired outcomes across all networks.

For elongated and narrow individual plastic greenhouses in the Group (3) extraction results, all networks exhibited certain omission errors. While UPerNet demonstrated severe adhesion phenomena, the remaining networks showed minor false extractions. For short and wide individual plastic greenhouses, common deep learning networks all exhibited both false positives and false negatives in extraction results, while only IPGENet showed minimal false positives.

Compared with these commonly used deep learning networks, IPGENet demonstrates superior performance in extracting both types of individual plastic greenhouses across diverse scenarios. This further validates IPGENet's outperformance over conventional deep learning networks under complex spatial configurations.

3.3. Extraction and Mapping of Individual Plastic Greenhouses in Shandong Province, China

Following the iteration of IPGENet, this study extracted Individual plastic greenhouses across Shandong Province. Partial extraction results from the training and validation phases are presented in Figure 8. During the extraction process, we observed that the existing sample dataset's categories failed to comprehensively cover all types of individual plastic greenhouses across the entire Shandong Province, resulting in suboptimal extraction outcomes that fell short of the anticipated performance. We propose a sample iteration method to enrich the sample dataset while ensuring efficiency and accuracy. First, regions with dense greenhouse distribution were identified in the remote sensing imagery. Representative individual plastic greenhouses were then selected from these areas and manually annotated to ensure labeling accuracy. With a sufficient number of labels paired with corresponding remote sensing images, models are trained using IPGENet for large-scale extraction of individual plastic greenhouses. Manual modification and annotation were performed on erroneous extraction results and omitted extraction results, respectively, while correct extractions were retained. This approach facilitates rapid expansion of dataset diversity and sample volume while maintaining label accuracy, substantially reducing manual time expenditure.

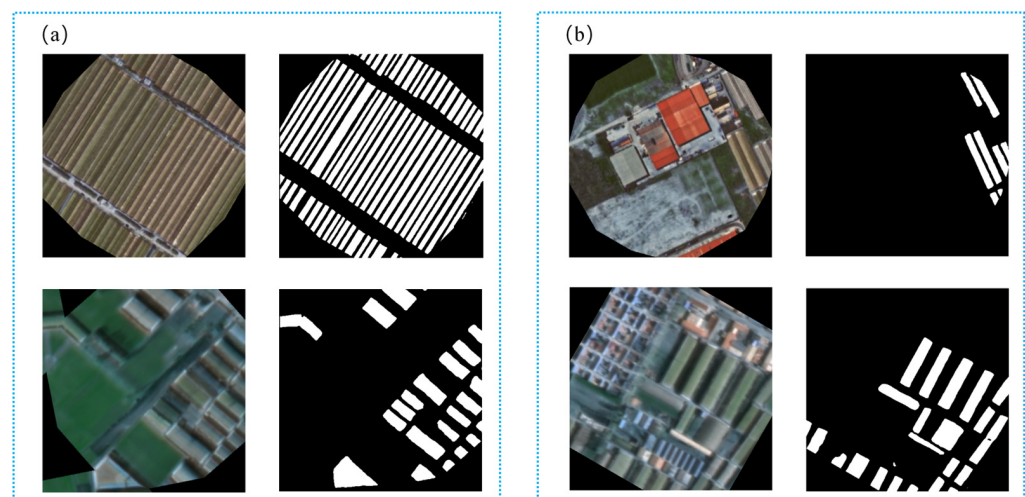


Figure 8. Partial extraction results during the model training and validation phase. (a) Remote sensing images and corresponding extraction results during the training phase. (b) Remote sensing images and corresponding extraction results during the validation phase.

The mapping results of individual plastic greenhouses across Shandong Province are presented in Figure 9. Spatial distribution analysis reveals a generally dispersed

pattern of greenhouses across the region. However, both the mapping results and area distribution density map (Figure 10) distinctly demonstrate that Weifang and Liaocheng cities exhibit the highest density concentrations, constituting the core areas of facility agriculture development in Shandong Province. The greenhouse distribution in Weihai, Zibo, and Dezhou was significantly lower than in other regions of Shandong Province, with insufficient large-scale development in facility agriculture, indicating relatively lagging modernization progress in this sector.

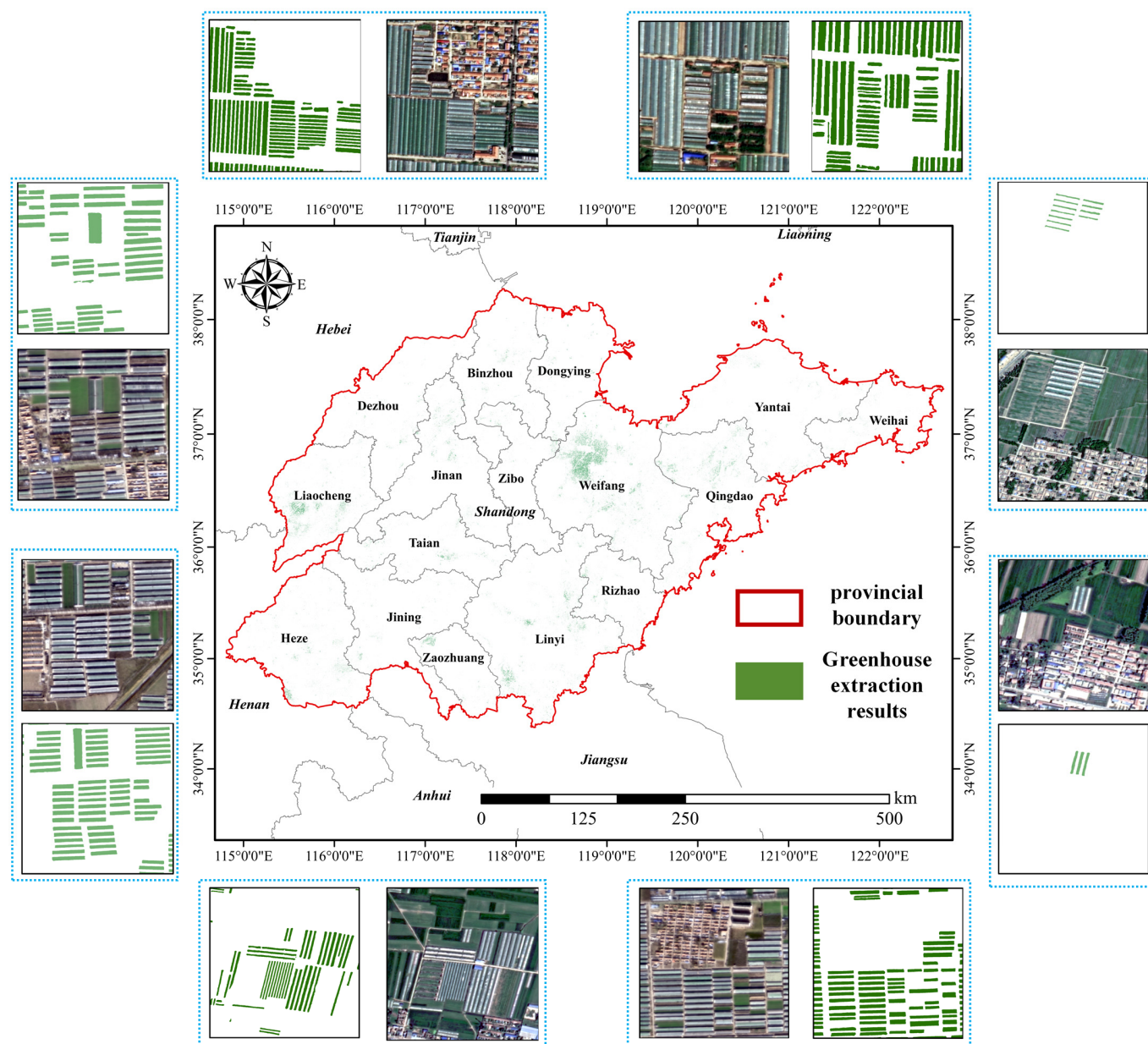


Figure 9. Cartographic presentation of extraction results for individual plastic greenhouses in Shandong Province, China.

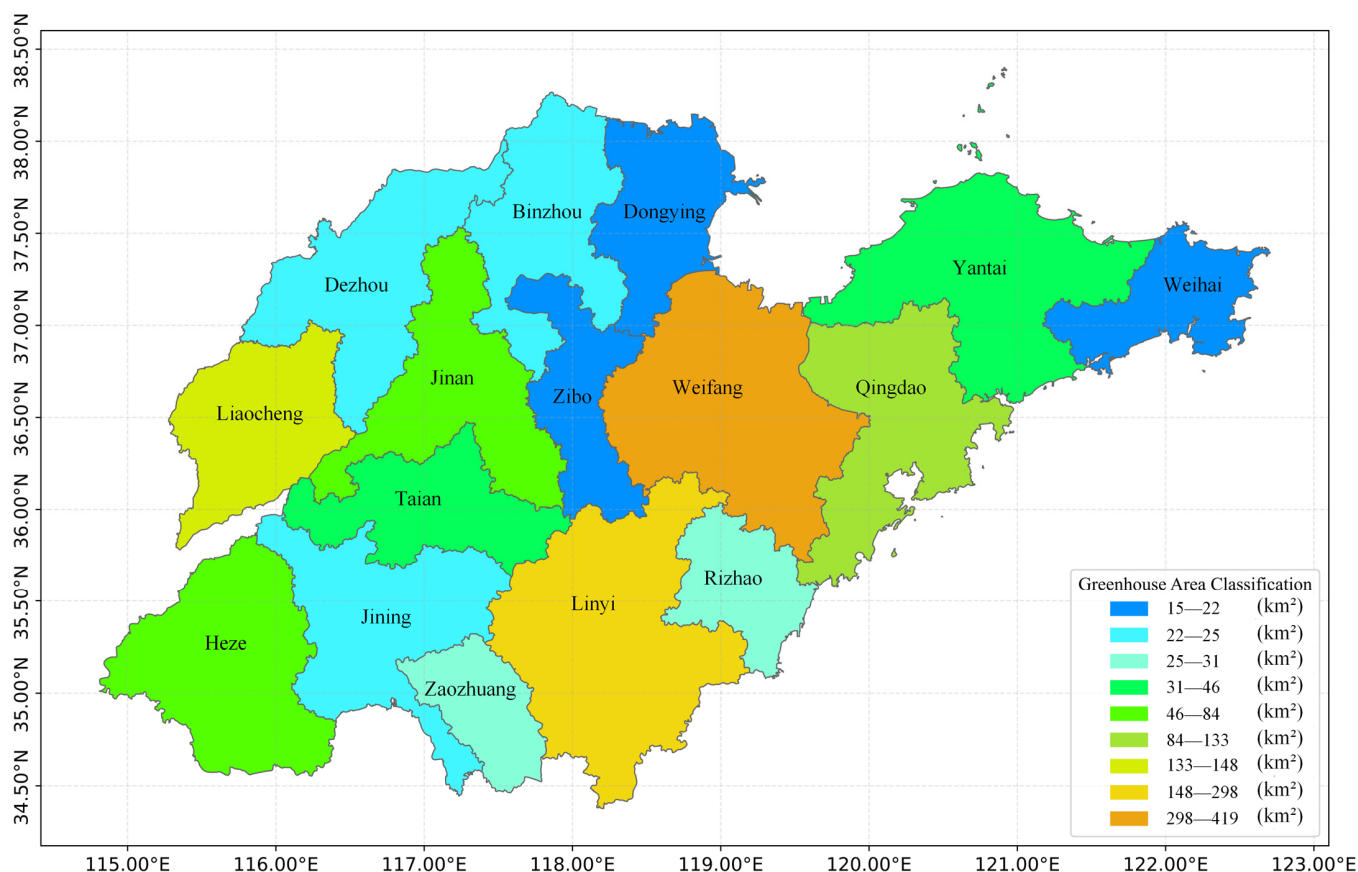


Figure 10. Area distribution density map of individual plastic greenhouses in prefecture-level cities of Shandong Province.

3.4. Statistical Analysis of Individual Plastic Greenhouses Across Shandong Province, China

The extracted quantity and area statistics of greenhouses in Shandong Province are presented in Table 4 and Figure 11. According to the data presented in Table 4, individual plastic greenhouse facilities in Shandong Province's 16 prefecture-level cities exhibit significant regional agglomeration characteristics. Based on the quantitative dimension of facility distribution, the total number of agricultural facilities across all prefectural cities in the province reaches 2.703 million units. Among these, three cities (Weifang (24.62%), Linyi (18.26%), and Liaocheng (9.41%)) collectively account for 52.29% of the provincial total, demonstrating a significant concentration of infrastructure. The spatial distribution exhibited a more pronounced concentration pattern, with Weifang City holding a commanding lead at 419.74 km² (34.85%), followed by Linyi (12.48%) and Liaocheng (11.75%). The cumulative proportion of the top three cities reached 59.08%. This “dominant leader accompanied by multiple strong regional clusters” distribution pattern indicates that the core driving areas of Shandong Province's modern facility agriculture development are primarily concentrated in central and southern Shandong regions. Particularly, Weifang City demonstrates its position as a large-scale facility agriculture cluster with intensive operations, evidenced by its substantial 24.62% numerical share and 34.85% areal proportion in provincial facility agriculture metrics.

Based on the analysis above, it is evident that facility agriculture development in Shandong Province exhibits both significant agglomeration advantages and faces severe challenges of regional imbalance. Future efforts must focus on promoting policy guidance and technology diffusion to facilitate the transition of provincial facility agriculture from a “unipolar breakthrough” model to “coordinated development.” This shift will fully leverage

modern agricultural facilities' functions in production stabilization, supply assurance, income growth, and rural prosperity.

Table 4. Statistical table of individual plastic greenhouse extraction results in cities of Shandong Province, China.

City	Number	Percentage of Total Quantity (%)	Area (km ²)	Percentage of Total Area (%)
Binzhou	55,465	2.05	22.45	1.86
Dezhou	46,954	1.74	23.65	1.96
Dongying	50,934	1.88	15.64	1.30
Heze	159,781	5.91	76.35	6.34
Jinan	100,655	3.72	47.76	3.97
Jining	67,747	2.51	24.34	2.02
Liaocheng	254,463	9.41	141.54	11.75
Linyi	493,633	18.26	150.26	12.48
Qingdao	234,979	8.69	110.75	9.20
Rizhao	74,380	2.75	27.35	2.27
Taian	93,620	3.46	35.47	2.94
Weifang	665,412	24.62	419.74	34.85
Weihai	67,263	2.49	16.36	1.36
Yantai	146,044	5.40	44.56	3.70
Zaozhuang	155,434	5.75	28.47	2.36
Zibo	36,683	1.36	19.74	1.64

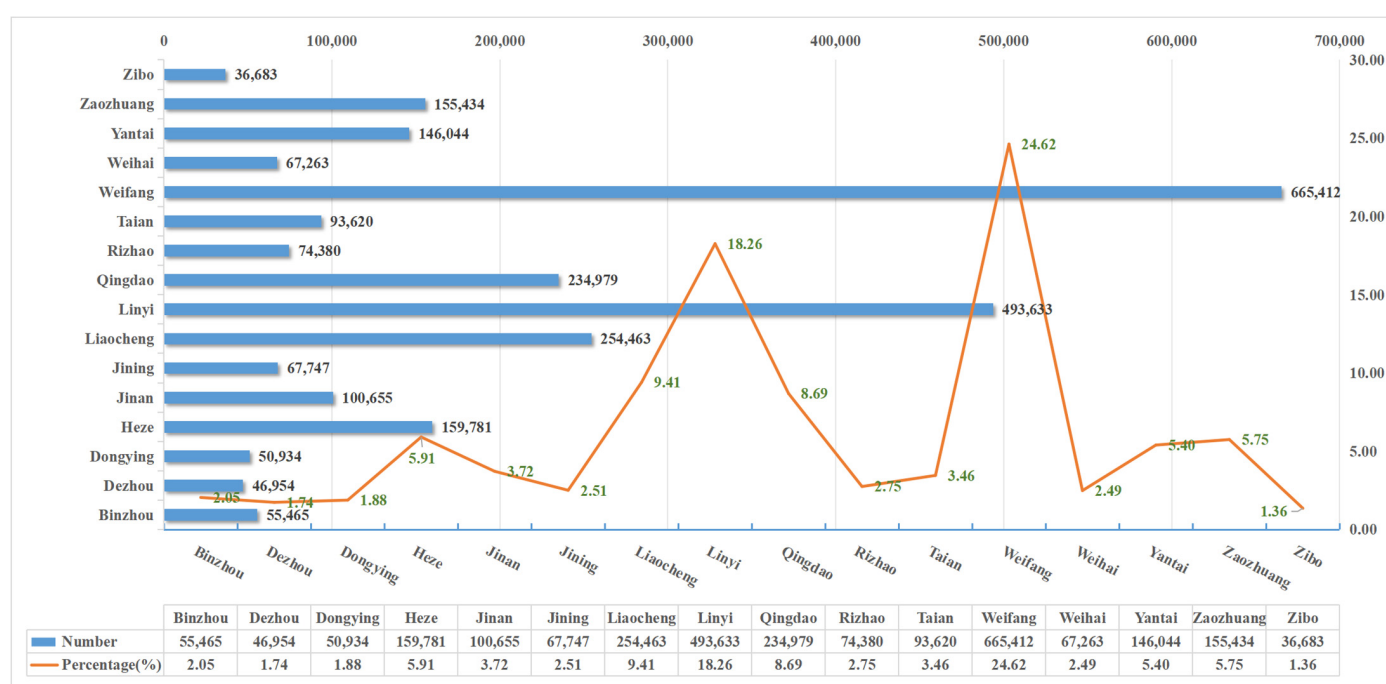


Figure 11. Statistical bar chart of individual plastic greenhouse quantity in prefecture-level cities of Shandong Province.

4. Discussion

4.1. Advantage

This study proposes a high-precision extraction network for large-scale individual plastic greenhouses. The method builds upon Swin-UNet as the foundational architecture and integrates a Multi-scale Feature Fusion Decoder (MFFD) to form the IPGENet model. Firstly, by hierarchically fusing multi-resolution feature maps through upsampling high-level semantic features and concatenating them with low-level detail features, followed by

convolutional processing, this approach enhances contextual awareness and enriches feature representation richness. Secondly, introducing a window-based attention mechanism inspired by Swin transformer principles efficiently captures local-to-global dependencies while maintaining computational efficiency. Finally, the iterative decoding process progressively refines outputs, integrating bilinear interpolation to ensure prediction-target alignment, which significantly enhances segmentation accuracy and detail restoration capabilities. The overall structure of IPGENet proposed in this study flexibly adapts to multi-scale inputs, balances efficiency and performance in complex scenarios, and enhances the model's recognition of edge and texture features of individual plastic greenhouses under such conditions.

4.2. Limitations and Future Directions

Although the proposed IPGENet in this study demonstrates excellent performance in extracting individual plastic greenhouses across large-scale regional areas in Shandong Province, there are still some limitations. First, this study failed to propose suitable post-processing methods for optimizing the extraction results of individual plastic greenhouses at large spatial scales. Second, the extraction results exhibited partial adhesion between adjacent greenhouses and omissions of visible structures, which may compromise the statistical accuracy of province-wide greenhouse quantification. Third, this study focused solely on the extraction of individual plastic greenhouses in Shandong Province, a plains region, and did not conduct large-scale extraction in regions with different geographic settings. The morphological characteristics of individual plastic greenhouse distributions may vary significantly across diverse geographical environments, constraining the model's generalizability. Furthermore, this study relied solely on GF-2 remote sensing imagery for extracting individual plastic greenhouses, exhibiting dependency on specific data that may compromise model generalization.

To address the above issues, this study will implement future improvements through two approaches: (1) designing post-processing methods applicable to large-scale individual plastic greenhouses in Shandong Province and other regions, optimizing extraction results to provide higher-precision outputs; (2) enhancing network architecture to improve edge recognition capability and segmentation accuracy for closely spaced greenhouses, achieving superior sub-meter-level extraction results. Enhanced model generalization capability was achieved by training with multi-regional datasets (e.g., mountainous areas) and multi-resolution remote sensing imagery (e.g., Sentinel and GF-7 imagery).

5. Conclusions

This study proposes a decoder module based on multi-scale feature fusion, integrates this module into the Swin-UNet architecture, and thereby designs the semantic segmentation framework IPGENet. By progressively fusing feature maps at different resolutions, the method upsamples high-level semantic information and concatenates it with low-level detail features. Subsequent convolutional processing enhances contextual awareness and enriches feature representation. Additionally, by employing a windowed attention mechanism, the model efficiently captures both local and global dependencies while reducing computational costs. To ensure sample accuracy while improving the efficiency of manual annotation, this study proposes an iterative sample approach that rapidly expands small-scale sample sets into large-scale datasets. IPGENet demonstrated superior performance in quantitative comparisons based on GF-2 remote sensing imagery of Shandong Province, achieving a recall, precision, Intersection over Union (IoU), and F1-score of 93.46%, 92.86%, 87.45%, and 93.16%, respectively. In addition to extracting single-span plastic greenhouses across Shandong Province, this study conducted quantitative and spatial

analyses on the quantity and area of greenhouses in each city within the province, revealing regional disparities in facility agriculture among Shandong's cities. Though the proposed IPGENet in this study demonstrates superior performance in extracting individual plastic greenhouses across large spatial scales in Shandong Province, it fails to develop a suitable post-processing method for the extraction results. Additionally, the outcomes exhibit issues of adhesion between adjacent greenhouse boundaries and false extractions. Future work will design a large-scale post-processing method tailored for individual plastic greenhouses in Shandong Province, further optimizing extraction results. Model generalization capability was enhanced by training with multi-resolution remote sensing imagery from diverse regions. The network architecture will be refined to enhance edge recognition capability and segmentation performance for elongated and narrow-shaped greenhouse structures, thereby advancing the application of deep learning and remote sensing technologies in sustainable agricultural development.

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

Funding: This research was supported by the Science and Disruptive Technology Program, Aerospace Information Research Institute, Chinese Academy of Sciences (No. E3Z219010F).

Data Availability Statement: Data are contained within the article.

Acknowledgments: We sincerely acknowledge the editor and all anonymous reviewers for their dedicated efforts and insightful comments, which have provided crucial guidance for future manuscript improvements.

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

Abbreviations

The following abbreviations are used in this manuscript:

MFFD	Multi-scale Feature Fusion Decoder
IPGENet	Individual Plastic Greenhouse Extraction Network
GF-2	Gaofen-2 satellite
DEM	Digital Elevation Model
SPP	Spatial Pyramid Pooling
Window CA	Window-Cross Attention
IoU	Intersection over Union
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative

References

1. Chen, Z.; Wu, Z.; Gao, J.; Cai, M.; Yang, X.; Chen, P.; Li, Q. A convolutional neural network for large-scale greenhouse extraction from satellite images considering spatial features. *Remote Sens.* **2022**, *14*, 4908. [[CrossRef](#)]
2. Guo, B.; Zhou, B.; Zhang, Z.; Li, K.; Wang, J.; Chen, J.; Papadakis, G. A critical review of the status of current greenhouse technology in China and development prospects. *Appl. Sci.* **2024**, *14*, 5952. [[CrossRef](#)]
3. Ma, H.; Feng, T.; Shen, X.; Luo, Z.; Chen, P.; Guan, B. Greenhouse extraction with high-resolution remote sensing imagery using fused fully convolutional network and object-oriented image analysis. *J. Appl. Remote Sens.* **2021**, *15*, 046502. [[CrossRef](#)]

4. Aguilar, M.Á.; Jiménez-Lao, R.; Nemmaoui, A.; Aguilar, F.J.; Koc-San, D.; Tarantino, E.; Chourak, M. Evaluation of the consistency of simultaneously acquired Sentinel-2 and Landsat 8 imagery on plastic covered greenhouses. *Remote Sens.* **2020**, *12*, 2015. [CrossRef]
5. van Delden, S.H.; SharathKumar, M.; Butturini, M.; Graamans, L.J.A.; Heuvelink, E.; Kacira, M.; Kaiser, E.; Klammer, R.S.; Klerkx, L.; Kootstra, G.; et al. Current status and future challenges in implementing and upscaling vertical farming systems. *Nat. Food* **2021**, *2*, 944–956. [CrossRef] [PubMed]
6. Novelli, A.; Aguilar, M.A.; Nemmaoui, A.; Aguilar, F.J.; Tarantino, E. Performance evaluation of object based greenhouse detection from Sentinel-2 MSI and Landsat 8 OLI data: A case study from Almería (Spain). *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *52*, 403–411. [CrossRef]
7. Picuno, P. Innovative material and improved technical design for a sustainable exploitation of agricultural plastic film. *Polym.-Plast. Technol. Eng.* **2014**, *53*, 1000–1011. Available online: <https://hdl.handle.net/11563/58958> (accessed on 5 July 2025). [CrossRef]
8. Picuno, P.; Sica, C.; Laviano, R.; Dimitrijević, A.; Scarascia-Mugnozza, G. Experimental tests and technical characteristics of regenerated films from agricultural plastics. *Polym. Degrad. Stab.* **2012**, *97*, 1654–1661. [CrossRef]
9. Sica, C.; Picuno, P. Spectro-radiometrical characterization of plastic nets for protected cultivation. In Proceedings of the International Symposium on High Technology for Greenhouse System Management: Greensys 2007, Naples, Italy, 4–6 October 2007; Volume 801, pp. 245–252. [CrossRef]
10. Picuno, P.; Tortora, A.; Capobianco, R.L. Analysis of plasticulture landscapes in Southern Italy through remote sensing and solid modelling techniques. *Landsc. Urban Plan.* **2011**, *100*, 45–56. [CrossRef]
11. Agüera, F.; Aguilar, F.J.; Aguilar, M.A. Using texture analysis to improve per-pixel classification of very high resolution images for mapping plastic greenhouses. *ISPRS J. Photogramm. Remote Sens.* **2008**, *63*, 635–646. [CrossRef]
12. Lu, L.; Di, L.; Ye, Y.A. decision-tree classifier for extracting transparent plastic-mulched landcover from Landsat-5 TM images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 4548–4558. [CrossRef]
13. Chi, M.; Plaza, A.; Benediktsson, J.A.; Sun, Z.; Shen, J.; Zhu, Y. Big data for remote sensing: Challenges and opportunities. *Proc. IEEE* **2016**, *104*, 2207–2219. [CrossRef]
14. Chen, W.; Xu, Y.; Zhang, Z.; Yang, L.; Pan, X.; Jia, Z. Mapping agricultural plastic greenhouses using Google Earth images and deep learning. *Comput. Electron. Agric.* **2021**, *191*, 106552. [CrossRef]
15. Yuan, J.; Wang, D.; Li, R. Remote sensing image segmentation by combining spectral and texture features. *IEEE Trans. Geosci. Remote Sens.* **2013**, *52*, 16–24. [CrossRef]
16. Sisodia, P.S.; Tiwari, V.; Kumar, A. Analysis of supervised maximum likelihood classification for remote sensing image. In Proceedings of the International Conference on Recent Advances and Innovations in Engineering (ICRAIE-2014), Jaipur, India, 9–11 May 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 1–4. [CrossRef]
17. Luo, K.; Zhang, H.; Zhu, C.; Jiao, T.; Samat, A.; Chen, Y.; Cheng, C. A novel method integrating sample migration and threshold optimization for high-precision greenhouse classification: Evidence from Southern China. *Geocarto Int.* **2025**, *40*, 2527308. [CrossRef]
18. Chen, S.; Chen, Y.; Gao, S.; Li, C.; Li, N.; Chen, L. A modified spectral remote sensing index to map plastic greenhouses in fragmented terrains. *Smart Agric. Technol.* **2025**, *11*, 100904. [CrossRef]
19. Sun, Y.; Zhang, Y.; Hao, J.; Li, J.; Ge, H.; Jiang, F.; Chen, F. Agricultural greenhouses datasets of 2010, 2016, and 2022 in China. *Sci. Data* **2025**, *12*, 1107. [CrossRef]
20. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef] [PubMed]
21. Li, J.; Huang, X.; Gong, J. Deep neural network for remote-sensing image interpretation: Status and perspectives. *Natl. Sci. Rev.* **2019**, *6*, 1082–1086. [CrossRef] [PubMed]
22. Cheng, G.; Yan, B.; Shi, P.; Li, K.; Yao, X.; Guo, L.; Han, J. Prototype-CNN for few-shot object detection in remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–10. [CrossRef]
23. Ding, J.; Xue, N.; Xia, G.-S.; Bai, X.; Yang, W.; Yang, M.Y.; Belongie, S.; Luo, J.; Datcu, M.; Pelillo, M.; et al. Object detection in aerial images: A large-scale benchmark and challenges. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 7778–7796. [CrossRef]
24. Ma, A.; Chen, D.; Zhong, Y.; Zheng, Z.; Zhang, L. National-scale greenhouse mapping for high spatial resolution remote sensing imagery using a dense object dual-task deep learning framework: A case study of China. *ISPRS J. Photogramm. Remote Sens.* **2021**, *181*, 279–294. [CrossRef]
25. Tong, X.; Zhang, X.; Fensholt, R.; Jensen, P.R.D.; Li, S.; Larsen, M.N.; Reiner, F.; Tian, F.; Brandt, M. Global area boom for greenhouse cultivation revealed by satellite mapping. *Nat. Food* **2024**, *5*, 513–523. [CrossRef] [PubMed]
26. Tian, X.; Chen, Z.; Li, Y.; Bai, Y. Crop classification in mountainous areas using object-oriented methods and multi-source data: A case study of Xishui county, China. *Agronomy* **2023**, *13*, 3037. [CrossRef]
27. Xie, J.; Tian, T.; Hu, R.; Yang, X.; Xu, Y.; Zan, L. A Novel Detector for Wind Turbines in Wide-Ranging, Multi-Scene Remote Sensing Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 17725–17738. [CrossRef]

28. Chen, D.; Zhong, Y.; Ma, A.; Cao, L. Dense greenhouse extraction in high spatial resolution remote sensing imagery. In Proceedings of the IGARSS 2020—2020 IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 26 September–2 October 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 4092–4095. [\[CrossRef\]](#)
29. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969. [\[CrossRef\]](#)
30. Wang, Q.; Chen, W.; Tang, H.; Pan, X.; Zhao, H.; Yang, B.; Gu, W. Simultaneous extracting area and quantity of agricultural greenhouses in large scale with deep learning method and high-resolution remote sensing images. *Sci. Total Environ.* **2023**, *872*, 162229. [\[CrossRef\]](#)
31. Liu, X.; Xiao, B.; Jiao, J.; Hong, R.; Li, Y.; Liu, P. Remote sensing detection and mapping of plastic greenhouses based on YOLOX+: A case study in Weifang, China. *Comput. Electron. Agric.* **2024**, *218*, 108702. [\[CrossRef\]](#)
32. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. Pytorch: An imperative style, high-performance deep learning library. In Proceedings of the 33rd International Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019; pp. 8026–8037.
33. Loshchilov, I.; Hutter, F. Decoupled weight decay regularization. *arXiv* **2017**, arXiv:1711.05101. [\[CrossRef\]](#)
34. Cao, H.; Wang, Y.; Chen, J.; Jiang, D.; Zhang, X.; Tian, Q.; Wang, M. Swin-unet: Unet-like pure transformer for medical image segmentation. In *European Conference on Computer Vision*; Springer Nature: Cham, Switzerland, 2022; pp. 205–218. [\[CrossRef\]](#)
35. Xiao, H.; Li, L.; Liu, Q.; Zhu, X.; Zhang, Q. Transformers in medical image segmentation: A review. *Biomed. Signal Process. Control* **2023**, *84*, 104791. [\[CrossRef\]](#)
36. Qin, J.; Huang, Y.; Wen, W. Multi-scale feature fusion residual network for single image super-resolution. *Neurocomputing* **2020**, *379*, 334–342. [\[CrossRef\]](#)
37. Zhang, C.; Chen, Y.; Yang, X.; Gao, S.; Li, F.; Kong, A.; Sun, L. Improved remote sensing image classification based on multi-scale feature fusion. *Remote Sens.* **2020**, *12*, 213. [\[CrossRef\]](#)
38. Wang, G.; Gan, X.; Cao, Q.; Zhai, Q. MFANet: Multi-scale feature fusion network with attention mechanism. *Vis. Comput.* **2023**, *39*, 2969–2980. [\[CrossRef\]](#)
39. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the 18th International Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015, Munich, Germany, 5–9 October 2015; Proceedings, Part III 18; Springer International Publishing: Cham, Switzerland, 2015; pp. 234–241. [\[CrossRef\]](#)
40. Xiao, T.; Liu, Y.; Zhou, B.; Jiang, Y.; Sun, J. Unified Perceptual Parsing for Scene Understanding. *arXiv* **2018**, arXiv:1807.10221. [\[CrossRef\]](#)
41. Li, H.; Xiong, P.; An, J.; Wang, L. Pyramid attention network for semantic segmentation. *arXiv* **2018**, arXiv:1805.10180. [\[CrossRef\]](#)
42. Zhang, Z.; Liu, Q.; Wang, Y. Road Extraction by Deep Residual U-Net. *arXiv* **2017**, arXiv:1711.10684. [\[CrossRef\]](#)

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.