RESEARCH PAPER

Integrating semantic segmentation and edge detection for agricultural greenhouse extraction

Yawen He, Feng Jin,* and Yongheng Li
China University of Petroleum (East China), College of Oceanography and Space Informatics, Qingdao, China

ABSTRACT. Agricultural greenhouses have a negative impact on the ecological environment while bringing huge economic and social benefits. Therefore, it is of great significance to obtain greenhouse information in a timely and accurate manner. Due to the complex spectral characteristics and dense spatial distribution characteristics of greenhouses, although the extraction of greenhouses based on a single semantic segmentation model can extract the area with high precision, the segmentation process has a serious problem of boundary adhesion between greenhouses, which makes it difficult to accurately obtain the quantity of greenhouses. To address this, our study proposes a method for greenhouse extraction that integrates semantic segmentation and edge constraints, using high-spatial-resolution remote sensing images to accurately extract the area and quantity of greenhouses. This method employs an improved semantic segmentation model (AtDy-D-LinkNet) to extract the greenhouse area, which embeds a convolutional attention module into the D-LinkNet and adopts a dynamic upsampling strategy, achieving precise greenhouse extraction. Experiments demonstrate that the improved model increased the recall, precision, $F1$ score, and intersection over union by 1.68%, 2.27%, 1.93%, and 3.54%, respectively, compared to the original model. To address the significant edge adhesion issue in semantic segmentation and accurately extract the quantity of greenhouses, we developed an edge constraint approach. This approach uses an edge detection model to extract greenhouse boundaries, further constrains the greenhouse surfaces, separates adhered greenhouses, and outputs vector patches representing individual greenhouses, thereby achieving precise greenhouse quantity extraction. The experiments show that this method effectively combines the advantages of semantic segmentation and edge detection. It not only ensures the accuracy of greenhouse area extraction but also effectively solves the boundary adhesion issue, significantly improving quantity extraction accuracy, resulting in vector patches that align with the actual area, quantity, and spatial distribution of greenhouses. This can provide a data foundation for greenhouse management and planning in agriculture.

1 Introduction

Food security is essential for human survival. However, with economic development and urban expansion, the amount of available arable land is decreasing, posing a threat to food security.[1,2] To address this issue, agricultural greenhouses (AGs) have been implemented as they can

*Address all correspondence to Feng Jin, z22160031@s.upc.edu.cn

effectively mitigate the impact of natural conditions on crop growth, extend the growing season, and enhance crop yields and land utilization rates.[3,4] Although AGs offer significant economic and social benefits, their large-scale irrigation and frequent fertilization practices can also contribute to the degradation of the natural environment. This includes issues such as soil salinization, soil acidification, eutrophication of water bodies, and the accumulation of soil pollutants.[5–9] Therefore, it is crucial to obtain timely and accurate information on the quantity, area, and spatial distribution of AGs. These data can provide support for intelligent monitoring and supervision of AGs, which is of great significance for preventing and controlling agricultural pollution and managing agricultural production.

The land survey method of on-site collection overly relies on human resources and is inefficient. The development of remote sensing technology, with its advantages of wider detection range, massive remote sensing data, and rapid information exchange, has been widely applied in land management, crop yield estimation, change monitoring, and other fields at home and abroad, becoming the mainstream land survey method. Since the 1960s, scholars have been studying the extraction of AGs information using remote sensing.[10–14] The traditional methods for AGs extraction are pixel-based classification and object-oriented classification. The pixel-based classification method uses spectral and texture features to classify pixels in the image. However, this method only considers individual pixels and does not consider context information, resulting in a "salt-and-pepper" effect in the classification. The object-oriented classification method addresses this issue by analyzing the shape, structure, texture, and spectral characteristics of ground objects and establishing classification rules. This method segments the image into objects with semantic consistency. However, it requires high segmentation parameters and thresholds and may not perform well in complex environments.

In recent years, there has been rapid development in artificial intelligence technology, particularly the widespread application of deep convolutional neural network models in the intelligent interpretation of remote sensing images.[15] These models have demonstrated outstanding performance and wide application prospects in various fields, such as land use and land cover classification,[16,17] ecosystem management,[18] and agricultural monitoring.[19] They have paved the way for innovative approaches to the precise extraction of AGs.[20–24] For instance, Yang et al.[20] compared the effectiveness of traditional SVM and classical semantic segmentation network in extracting trellis structures from multispectral and ultrahigh resolution unmanned aerial vehicle (UAV) data. Their results showed that the semantic segmentation model had higher accuracy and efficiency. Similarly, Chen et al.[21] combined convolutional neural network and long- and short-term memory network to develop a spatial long- and short-term memory structure, which improved the accuracy of AG boundary extraction on a large scale.

Currently, the semantic segmentation methods for extracting AGs from medium to high spatial resolution remote sensing images yield unsatisfactory results, with coarse predictions. This can be attributed to two main factors: (1) the fine pixels captured by high spatial resolution images introduce more noise due to factors, such as shadows, light reflections, and occlusions, exacerbating the intraclass diversity within greenhouses. Consequently, classical semantic segmentation models encounter issues, such as fragmentation within greenhouses and blurring at their edges. (2) Greenhouses exhibit dense spatial distribution characteristics. From a macroscopic perspective, greenhouses are often densely constructed, with some areas characterized by extremely dense spatial distribution. Even with high-resolution imagery, it is challenging for semantic segmentation models to accurately identify individual greenhouses.[25] Instead, they tend to recognize contiguous clusters of greenhouses as a single entity, leading to difficulties in accurately extracting the area and number of greenhouses.

To mitigate the influence of edge adhesion, researchers have utilized high-resolution UAV imagery with a resolution of 0.1 m to extract greenhouses, yielding highly detailed interpretation results. However, compared to satellite imagery, UAV imagery acquisition entails higher costs and covers smaller areas, making it difficult to achieve large-scale greenhouse extraction.[22] Therefore, enhancing the model's perception of greenhouse edges is crucial in current research on fine-scale greenhouse extraction based on satellite imagery. Li et al.[23] used ResNet as the encoder to create an end-to-end semantic segmentation model called EAGNet. This model effectively preserved boundary information using the edge attention mechanism, identifying densely distributed single AGs. However, the excessive focus on edge separation has led to a significant

disparity between the extracted gaps and the actual situation, affecting the accuracy of AG area extraction. Additionally, the complex and extensive network structure hinders operational efficiency. Zhang et al.[24] utilized HRNetV2[26] as the backbone network and developed edge refinement modules to enhance HBRNet, which showed superior performance compared to the classical model. However, solely relying on a single semantic segmentation model cannot fundamentally address edge adhesion, making it challenging to accurately extract information regarding the number of greenhouses. Therefore, some scholars have adopted a dual-branch structure, combining semantic segmentation for area extraction and object detection for quantity extraction, to extract greenhouse information in a coordinated manner using two models,[25,27] achieving precise greenhouse extraction. However, in this approach, the predictions of the two models are mutually independent, and they have not truly achieved integration in terms of prediction result scale.

In response to the aforementioned issues, the main contributions of this study are as follows: (1) addressing the issues of fragmentation and edge blurring of AG extraction in high spatial resolution imagery by employing the D-LinkNet architecture embedded with convolutional block attention module (CBAM) and replacing traditional transpose convolution with a dynamic upsampling strategy, proposing the AtDy-D-LinkNet model. This model enhances the learning and recognition capabilities of the greenhouse spectra and spatial features, significantly improving the accuracy of greenhouse edge recognition. (2) In response to the issue of edge adhesion in semantic segmentation arising from the dense spatial distribution of AGs and the limited spatial resolution of remote sensing imagery, this study proposes a "integrating semantic segmentation and edge detection for greenhouse extraction." This method utilizes an "edge constraint" by employing an edge detection model to extract greenhouse boundaries in addition to the surface extraction of greenhouses by AtDy-D-LinkNet. At the result scale, it leverages greenhouse edge constraints to separate aggregated greenhouses, obtaining regular greenhouse vector patches. This ultimately achieves refined extraction of greenhouse area, quantity, and spatial distribution in the study area.

2 Data

2.1 Study Area

Shouguang City is situated in the northwest of Weifang City and the southwest coast of Laizhou Bay on the Bohai Sea. It is a county-level city under the jurisdiction of Shandong Province, covering a total area of about 2072 km^2, as depicted in Fig. 1. Shouguang City is located between $36°41'N \sim 37°19'N$, $118°32'E \sim 119°10'E$, belonging to the continental climate of the warm temperate monsoon zone, with the characteristics of cold winters and hot summers, rain, and heat. Shouguang City covers a total area of about 2.06 million acres of cultivated crops, with the vegetable planting area reaching up to 623,000 acres and an annual output of 3.798 million tons, earning it the nickname "China's vegetable basket." This typical spatial distribution makes it an ideal area for remote sensing extraction of AGs.

2.2 Data Collection and Labeling

In this study, the submeter image from the Chinese commercial satellite SuperView-1 in October 2021 was utilized as the experimental data. The spatial resolution of the image after preprocessing, including geometric correction, orthorectification, fusion, and mosaic, was 0.5 m. The image contains four wavebands: blue, green, red, and near infrared, with only the visible band being utilized in this study.

The quality of the sample directly affects the experimental results. To ensure the reliability of the experiment, this study utilized a random scattering and manual screening method to uniformly select image blocks sized 1000×1000 as sample images. The selected samples covered all types of AGs in the study area and included an adequate number of pure background samples to enhance the robustness and generalization of the model. We annotated the greenhouse texture labels and edge labels separately based on the different training characteristics and sample requirements of the semantic segmentation model and edge detection model. Through augmentation techniques, such as rotation, mirroring, color transformation, and random cropping, we expanded the sample set to obtain 1600 examples of greenhouse texture samples and greenhouse
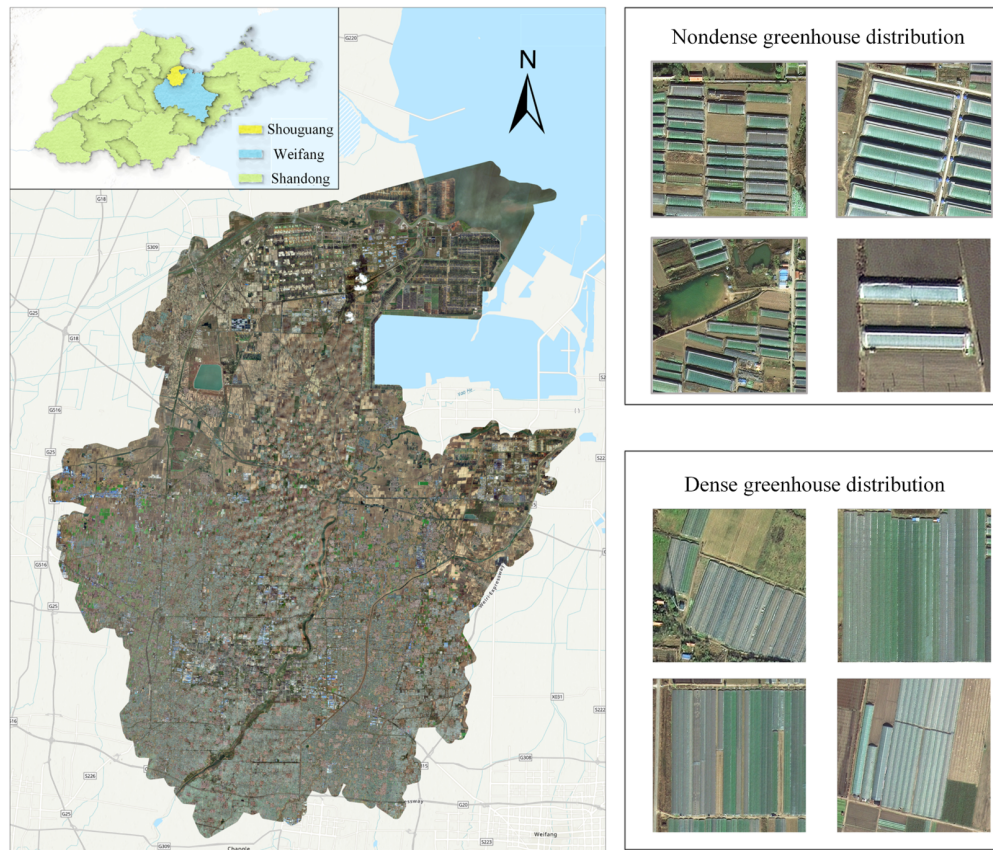
Fig. 1 Schematic diagram of the study area.

edge samples, each with a size of 640×640. These samples were then randomly divided into training, validation, and testing sets in an 8:1:1 ratio.

3 Methodology

3.1 Overview of the Methodology

To achieve high-precision extraction of greenhouse area, quantity, and spatial distribution information, this study proposes integrating semantic segmentation and edge detection for greenhouse extraction. The technical route of this method is illustrated in Fig. 2 and mainly consists of three parts: data preprocessing module, model training module, and greenhouse extraction module.

In the data preprocessing and model training modules, the previously described methods are utilized to construct greenhouse texture and edge datasets, which are then employed to train the semantic segmentation and edge detection models, respectively, to obtain optimal model parameters. In the greenhouse extraction module, the process involves several steps. First, the greenhouse texture surface is extracted based on the semantic segmentation model to obtain preliminary results, where multiple densely distributed greenhouses may be identified as a whole. Then the significant visual boundaries of the greenhouses are extracted based on the edge detection model, accurately identifying even the edges of densely distributed greenhouses. Finally, through postprocessing algorithms, semantic segmentation and edge detection are fused at the result scale to ultimately output refined greenhouse vector polygons.

3.2 AtDy-D-LinkNet

The input remote sensing image is represented as $F \in R^{C \times H \times W}$, where C is the number of channels, and $H \times W$ denotes the spatial resolution of the image. The objective of this study is to automatically segment remote sensing images and generate pixel-level semantic feature maps of size $H \times W$, aiming to accurately extracting AGs. The proposed AtDy-D-LinkNet is constructed based on the standard encoder–decoder U-shaped architecture, as illustrated in Fig. 3.
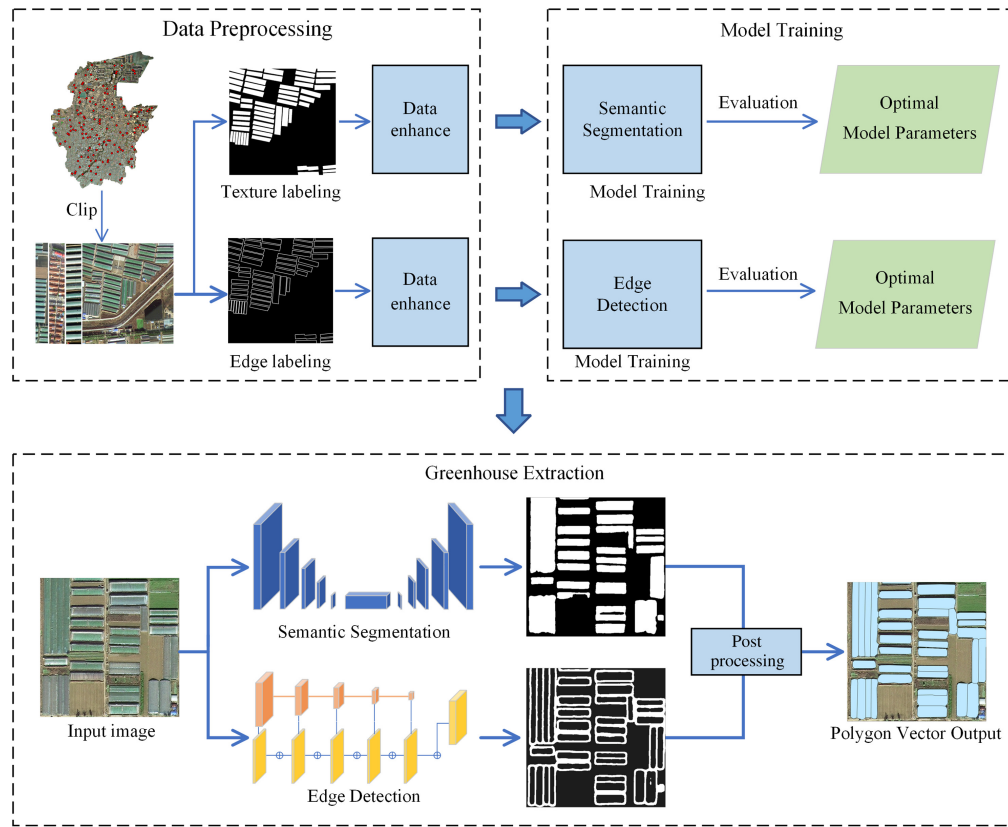
Fig. 2 Technical route for integrating semantic segmentation and edge detection for greenhouse extraction.
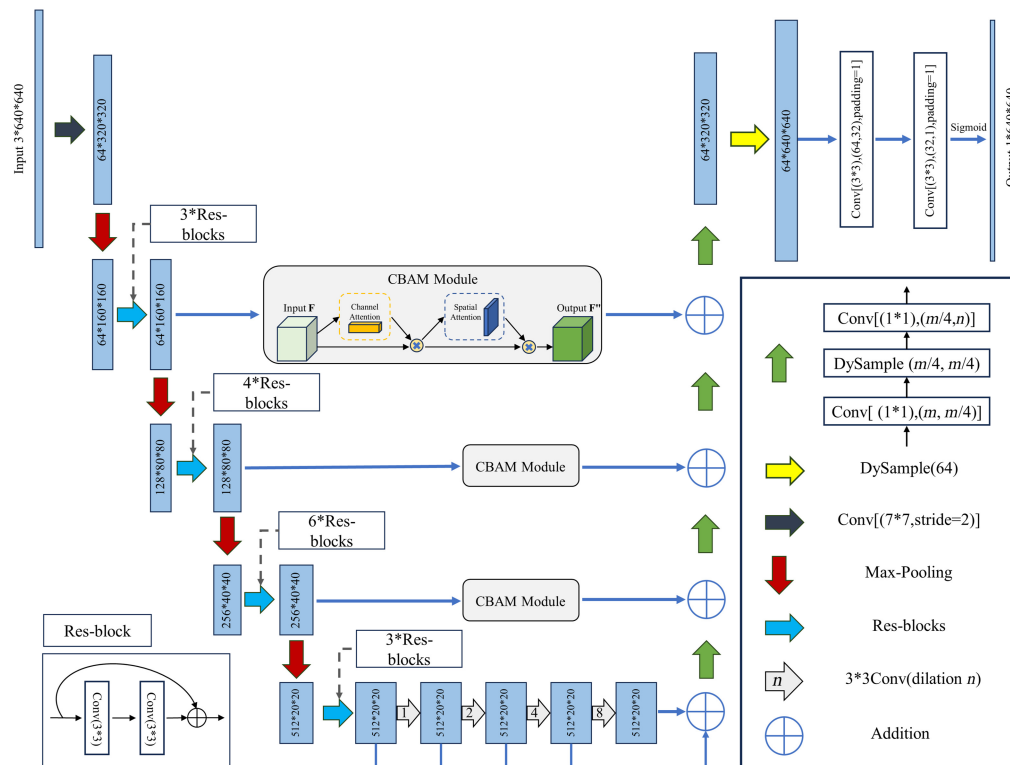

Fig. 3 AtDy-D-LinkNet module structure diagram.

To overcome the limitations mentioned in the first section, AtDy-D-LinkNet utilizes spatial and channel-wise complementary convolutional attention modules,[28] along with a dynamic sampling (DySample) upsampling strategy,[29] to further enhance the semantic segmentation quality of remote sensing images. Compared to the standard D-LinkNet, AtDy-D-LinkNet incorporates the CBAM into the multiscale skip connection process. It leverages the advantages of channel attention and spatial attention mechanisms to perceive and aggregate semantic information from distant contexts, thereby enhancing the expressive power of the model's semantic features. Additionally, DySample is employed in AtDy-D-LinkNet to replace the original transpose convolution for upsampling, effectively acquiring and expanding the semantic information of deep feature maps, thereby enhancing the model's feature extraction performance.

3.2.1 *CBAM attention module*

The original D-LinkNet adopts skip connections to merge the feature maps of the encoder and decoder, aiming to address the issue of spatial and channel information loss caused by the downsampling process. However, due to the significant semantic gap between the connected convolutional feature maps, the direct addition fusion method is prone to loss of image smoothness and the introduction of pseudoboundaries.[30] Therefore, the proposed AtDy-D-LinkNet introduces the CBAM attention module at the skip connection between the encoder and decoder networks, as illustrated in Fig. 4. This module concatenates two independent attention mechanisms: channel attention and spatial attention. Serving as a bridge, CBAM enhances the semantic information of the encoding feature maps through two layers of convolutional attention, making the skip connections smoother and further strengthening the model's ability to capture long-distance channel and spatial information.

The CBAM attention module concatenates channel attention and spatial attention. Given an input image, the two attention modules compute complementary attention focusing on "what" and "where" aspects, which can be mathematically represented by Eq. (1). Specifically, the feature maps F sampled and output by the Res-Blocks in the encoder are first passed through the channel attention module of CBAM. This module initially employs max-pooling and average-pooling to aggregate spatial feature information, generating two tensors describing different spatial contexts. These tensors then enter a perceptron with a hidden layer to share the internal channel parameters of the tensors. Subsequently, the module outputs channel attention maps separately, which are then combined to perceive channel information elementwise and obtain channel attention weights M_c through Sigmoid activation, as shown in Eq. (2). Finally, the refined intermediate layer F is obtained by multiplying M_c with the input F'.

Spatial attention primarily focuses on "where." F' serves as the input to the spatial attention module, where it undergoes both max-pooling and average-pooling operations. Subsequently,
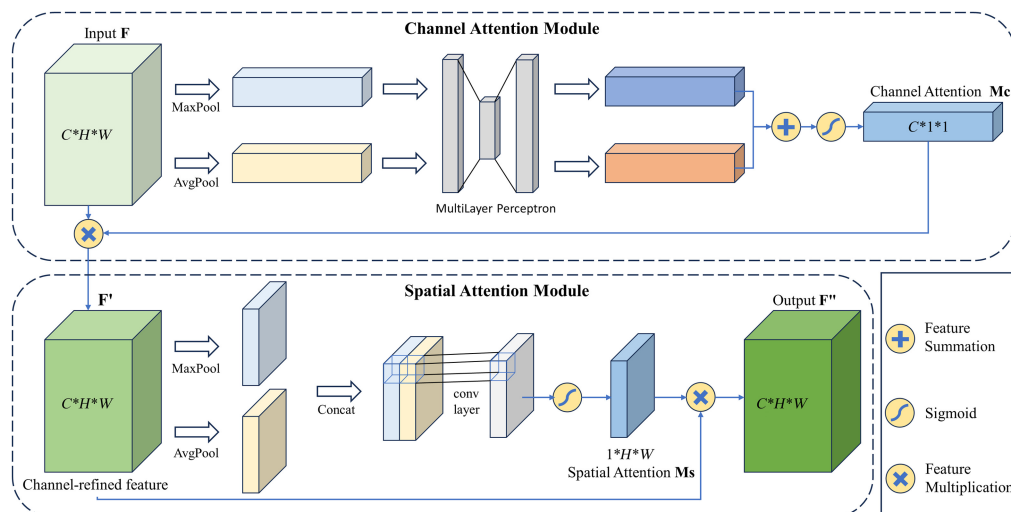

Fig. 4 CBAM module structure diagram.

the resulting feature maps are concatenated along the channel dimension and passed through convolutional layers to obtain an effective feature descriptor tensor. Afterward, the Sigmoid activation is applied to obtain spatial attention weights M_s, as shown in Eq. (3). Finally, F'' is obtained by multiplying M_s with the intermediate layer F':

$$F' = M_c(F) \otimes F,$$
$$F'' = M_s(F') \otimes F', \tag{1}$$

$$M_c(F) = \sigma(\text{MLP}(\text{AVgPool}(F)) + \text{MLP}(\text{MaxPool}(F)))$$
$$= \sigma(W_1(W_0(F^c_{\text{avg}})) + W_1(W_0(F^c_{\text{max}}))), \tag{2}$$

$$M_s(F) = \sigma(f^{7 \times 7}([\text{AVgPool}(F); \text{MaxPool}(F)]))$$
$$= \sigma(f^{7 \times 7}([F^s_{\text{avg}}; F^s_{\text{max}}])). \tag{3}$$

3.2.2 DySample upsampling strategy

In the field of image segmentation, upsampling is a crucial technique used to increase the resolution of low-resolution images or feature maps to higher resolutions, aiming to enhance image details or recover lost information. Upsampling methods mainly include linear interpolation-based upsampling and deep learning-based upsampling. Linear interpolation-based upsampling methods, such as nearest neighbor interpolation and bilinear interpolation, are simple and easy to use. However, they interpolate low-resolution features based on fixed rules, ignoring the semantic information in the feature maps, leading to information loss. Deep learning-based upsampling fundamentally involves enlarging the size of feature maps by training transpose convolution kernels. The original D-LinkNet adopts transpose convolution for upsampling, which learns features by sharing parameters, effectively utilizing network parameters. However, this approach is prone to causing the checkerboard artifact and information loss.

To overcome the limitations of traditional methods, AtDy-D-LinkNet introduces DySample to replace the original transpose convolution. DySample is an innovative dynamic upsampling method designed to enhance the model's capacity to capture spatial relationships and improve segmentation accuracy. It considers upsampling as point resampling, accepting input feature maps and generating corresponding high-resolution feature maps as output. It is worth noting that DySample can produce three variants (DySample+, DySample-S, and DySample-S+) with different internal structures or parameter selections. Experiments were conducted on the greenhouse dataset to test these four variants, and based on a comprehensive evaluation of accuracy and effectiveness, the basic DySample was selected as the upsampling module for AtDy-D-LinkNet.

The structure of the basic DySample is illustrated in Fig. 5. Initially, the input image X dynamically generates initial offsets used to adjust the sampling positions in the feature map through convolution. Subsequently, the offset O is obtained through pixel shuffle and combined with the original grid of the input image to dynamically calculate the sampling point grid S.
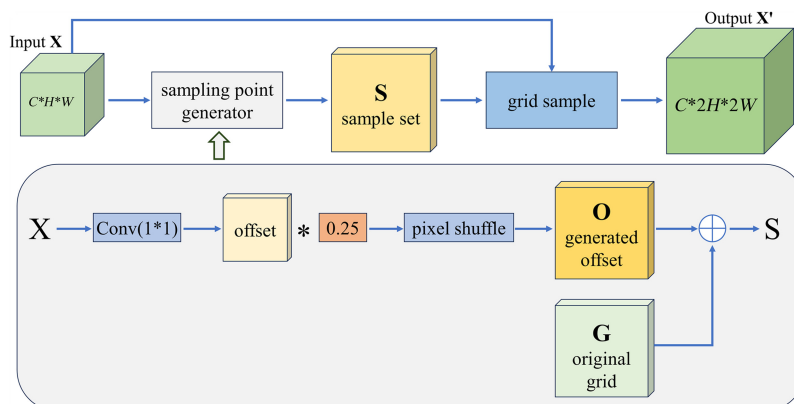

Fig. 5 DySample module structure diagram.

Finally, grid sampling is performed based on bilinear interpolation. By dynamically generating, learning, and optimizing sampling positions, DySample enables the model to better preserve spatial information between features, thus avoiding information loss and blurring phenomena. This enhancement leads to significant performance improvements in tasks, such as semantic segmentation, with higher efficiency and accuracy compared to traditional methods.

3.3 Edge Constraint Method

The edge constraint method proposed in this paper is based on the high-precision extraction of greenhouse surfaces. First, the edge detection model's sensitivity to abrupt changes in greenhouse texture and spectral features is leveraged to accurately detect the edges of the greenhouse, obtaining edge data as auxiliary information. Then in the postprocessing stage, the edge data are used to constrain and optimize the surface results, ensuring the accuracy of the edges while maintaining overall consistency, and ultimately integrating to generate vector polygons of the greenhouses, achieving a precise extraction of the greenhouses.

3.3.1 *Edge detection model*

The performance of the edge detection model directly determines the quality of the results. Traditional edge detection algorithms[31–34] quantize local image features to explore object boundaries. However, they lack a global perspective and rely solely on low-level surface features, making it difficult to express high-level semantic boundaries of targets. Deep learning-based edge detection methods address these limitations by essentially taking a classification approach. The objective is to categorize object edges and background into two distinct categories to achieve edge extraction. HED[35] first achieved end-to-end edge detection. It is based on the VGG16 framework and introduces a deep supervision mechanism, consisting of a backbone network and five side output branches at different levels. It adopts a holistic-nested structure to extract semantic information of varying depths from images at different scales, followed by upsampling to restore image resolution and outputting. Finally, the feature maps output by each branch are cascaded and fused to obtain the edge detection results. DexiNed,[36] inspired by HED and Xception,[37] combines the holistic-nested, side output edge detection framework with the Xception backbone structure to reduce the loss of edge features during network downsampling. It uses transposed convolution for upsampling, refining the prediction of target edges. Therefore, in this study, DexiNed is used as the foundational model for the edge constraint method.

3.3.2 *Postprocessing method*

Postprocessing is crucial for integrating the results of semantic segmentation and edge detection. This method first enhances the accuracy and reliability of the model's extraction results through mathematical morphology operations. Then, using a raster-to-vector conversion algorithm, the extracted surfaces and edges of the greenhouse are transformed into line vector features, which are easier to process and edit. Subsequently, through vector merging, the fine edges of the greenhouse constrain the greenhouse surfaces, ensuring clear boundaries between greenhouses and separating adhered greenhouses. Finally, through comprehensive processing of vector lines and polygons, the spatial data of the greenhouses are transformed and optimized, leading to precise extraction of the greenhouses. The main technical process of postprocessing is illustrated in Fig. 6 and detailed as follows.

(1) In the postprocessing of semantic segmentation, initially, a threshold of 0.5 is applied to binarize the extracted greenhouse surface results and fill holes smaller than 500 pixels. Subsequently, raster-to-polygon and polygon-to-line conversions are conducted to obtain greenhouse line features. Finally, the Douglas–Peucker algorithm is employed to retain the crucial bends in the lines, achieving line simplification.

(2) In the postprocessing of edge detection, the extracted greenhouse edge results are binarized using an optimal threshold based on an automatic image thresholding algorithm. Next, a morphology-based edge thinning algorithm is used to process the binary raster, resulting in a one-pixel-width greenhouse edge raster dataset. Following the same approach, raster-to-line conversion and line simplification are performed.
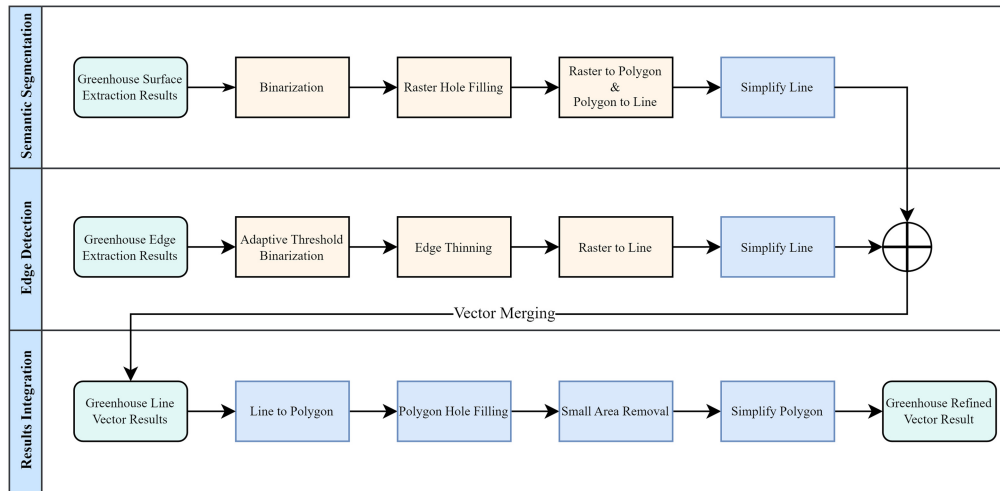
Fig. 6 Result postprocessing flowchart.

(3) In the results integration phase, line vector merging is initially conducted to constrain the surface segmentation results with the greenhouse edge detection outcomes. Subsequently, line-to-polygon conversion is employed to obtain integrated boundary-constrained greenhouse patches. To refine the results, holes in the greenhouse patches smaller than 300 square units are filled, and fragmented patches smaller than 300 square units are removed. Finally, the Douglas–Peucker algorithm is utilized to simplify the polygon features, yielding in refined greenhouse vector results.

3.4 Model Evaluation Methods

To quantitatively evaluate the extraction performance of the AtDy-D-LinkNet model, four mainstream evaluation metrics were selected in the experiments: recall, precision, F_1 score, and intersection over union (IoU). Recall describes the proportion of correctly predicted pixels among all actual greenhouse pixels, as shown in Eq. (4). Precision represents the proportion of correctly predicted greenhouse pixels among all pixels predicted as greenhouses by the model, as shown in Eq. (5); F_1 score comprehensively considers recall and precision by computing their weighted average, as shown in Eq. (6). IoU measures the overlap between the extraction results and the ground truth, as depicted in Eq. (7):

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{4}$$

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{5}$$

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}, \tag{6}$$

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}. \tag{7}$$

In the above formulas, true positive (TP) represents the number of samples correctly predicted as positive by the model, true negative (TN) represents the number of samples correctly predicted as negative by the model, false positive (FP) represents the number of samples incorrectly predicted as positive by the model, and false negative (FN) represents the number of samples incorrectly predicted as negative by the model.

Due to the test set's inability to fully cover the complex conditions in the study area, it is challenging to comprehensively assess the model's actual performance in extracting greenhouse area from large-scale images based solely on accuracy data from the test set. Therefore, experiments were conducted in three typical areas (zone 1, zone 2, and zone 3) within Shouguang City, each consisting of 3722×3722 pixel patches. These areas encompass various types of greenhouses with different spatial distributions, situated amidst complex background features.

They were utilized to evaluate the accuracy of greenhouse area and quantity extraction in large-scale images.

To validate the effectiveness of this method in greenhouse area extraction, experiments were conducted by computing the confusion matrix between the ground truth images and the predicted images in typical areas. Two metrics, F_1 score and kappa coefficient, were calculated to comprehensively measure the accuracy of the extraction. Additionally, the area accuracy (AA) was calculated to directly assess the difference between the extracted greenhouse area (A_{pre}) and the actual greenhouse area (A_{gt}), as shown in Eq. (8). The kappa coefficient is utilized to measure classification consistency, considering both the model's prediction consistency and the random chance of classification, as described in Eq. (9), where p_o represents the classification accuracy, and p_e denotes the expected accuracy. For evaluating the effectiveness of greenhouse quantity extraction, experiments involved counting the number of vector patches output by the edge constraint module to obtain the extracted greenhouse quantity (N_{pre}). This, combined with the manually annotated actual greenhouse quantity (N_{gt}), allowed for the calculation of quantity accuracy (QA). QA describes the difference between the predicted and actual greenhouse counts, providing an intuitive reflection of the effectiveness of greenhouse quantity extraction, as shown in Eq. (10):

$$AA = 1 - \frac{|A_{pre} - A_{gt}|}{A_{gt}}, \tag{8}$$

$$k = \frac{p_o - p_e}{1 - p_e},$$
$$p_o = \text{accuracy},$$
$$p_e = \frac{(TP + FN) \times (TP + FP) + (TN + FP) \times (TN + FN)}{(TP + FP + TN + FN)^2}, \tag{9}$$

$$QA = 1 - \frac{|N_{pre} - N_{gt}|}{N_{gt}}. \tag{10}$$

4 Results and Analysis

4.1 Semantic Segmentation Model Evaluation

To quantitatively evaluate the performance of the proposed model, this study compared the extraction accuracy of the AtDy-D-LinkNet model with four representative semantic segmentation models (FCN8S, UNet, DeepLabV3, and D-LinkNet) based on the same environment using the test set. The results in Table 1 demonstrate that the accuracy of the AtDy-D-LinkNet model proposed in this paper is significantly superior to other models. The extraction accuracy of DeepLabV3, FCN8S, and D-LinkNet is similar. The AtDy-D-LinkNet model achieves 0.9083, 0.9107, 0.9081, and 0.8634 for recall, precision, F_1 score, and IoU, respectively. Compared to the original D-LinkNet, AtDy-D-LinkNet exhibits improvements of 1.68%, 2.27%, 1.93%, and 3.54% in these metrics, indicating high segmentation performance.

Table 1 Accuracy evaluation of semantic segmentation models.

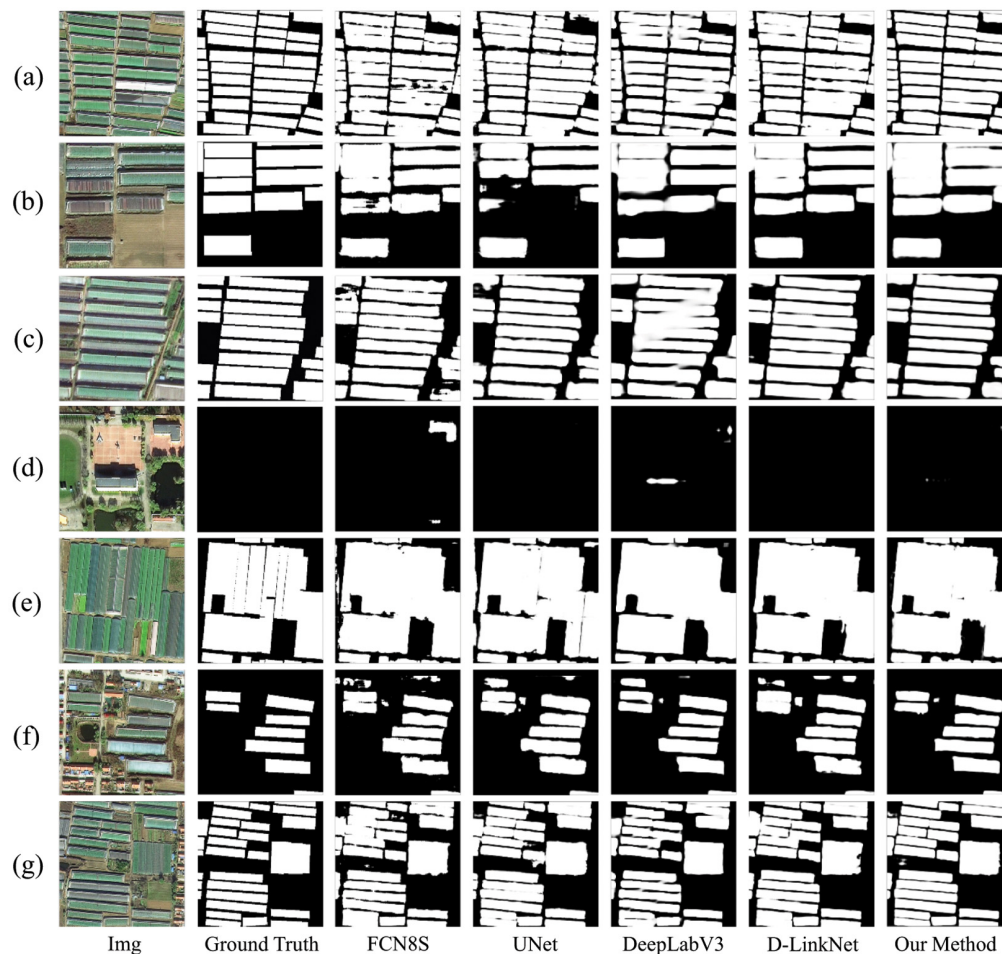| Model	Precision	Recall	F_1	IoU
FCN8S	0.8793	0.8743	0.8759	0.8143
UNet	0.8665	0.8617	0.8635	0.8088
DeepLabV3	0.8962	0.8923	0.8940	0.8475
D-LinkNet	0.8932	0.8904	0.8909	0.8340
AtDy-D-LinkNet	0.9083	0.9107	0.9081	0.8634

Fig. 7 Comparison of semantic segmentation model predictions. The figure compares the greenhouse extraction results of the ground truth and five semantic segmentation models for various regions (a)–(g) within the study area.

Figure 7 illustrates the extraction results of each model. Overall, AtDy-D-LinkNet extracts greenhouses more closely to reality. The coupling use of channel and spatial attention modules along with dynamic upsampling strategy yields excellent results: (1) this model significantly reduces the presence of holes and fragments on the greenhouse surface while improving the accuracy of boundary recognition between greenhouses and background objects, facilitating the extraction of more complete and regular greenhouses. (2) The model reduces misidentification of background objects, such as buildings, roads, and farmland, demonstrating stronger robustness in complex scenes. (3) The model exhibits a stronger capability in greenhouse boundary recognition and separation, effectively alleviating edge adhesion issues caused by dense greenhouse distributions, as shown in Fig. 7(a). However, in Fig. 7(e), where the distribution of greenhouses is extremely dense, semantic segmentation alone cannot extract individual greenhouses, resulting in multiple greenhouses being identified as a single entity.

4.2 Evaluation of Greenhouse Area Accuracy Using AtDy-D-LinkNet

To evaluate the practical performance of the AtDy-D-LinkNet model in extracting greenhouse areas, we calculated the confusion matrix between the extraction results of AtDy-D-LinkNet and the ground truth annotation images, obtaining comprehensive accuracy metrics F_1 and kappa, and directly calculate the AA for greenhouse extraction. As shown in Table 2, AtDy-D-LinkNet achieved high accuracy in extracting greenhouse areas in the three typical regions, indicating its high overall quality in identifying large-scale image features.

It is worth noting that the models overestimate the areas in zone 1 and zone 2, while underestimating the areas in zone 3. This phenomenon arises due to the different spatial distributions of

Table 2 Area extraction accuracy of typical zone.

Typical zones	A_{gt} (m^2)	A_{pre} (m^2)	AA	F_1	Kappa
Zone 1	1423130.25	1444843.00	0.9847	0.9549	0.8910
Zone 2	1497838.75	1488383.50	0.9937	0.9382	0.8568
Zone 3	1605095.25	1480240.75	0.9222	0.9216	0.8243

greenhouses in the three typical areas. In zone 1 and zone 2, the greenhouses are densely distributed, which causes the greenhouses to be closely positioned, leading to edge adhesion during segmentation and misclassification of some background pixels, resulting in an overestimation of the area. In zone 3, the greenhouses are mainly distributed discretely. The attention mechanism and dynamic sampling strategy in the model refine the edges of the greenhouses, causing the extracted greenhouse boundaries to contract inward. Although this refinement improves the issue of edge adhesion between greenhouses, it also leads to an underestimation of the extracted area compared to the actual area.

4.3 Application of Edge Constraint Method

Figure 8 demonstrates the practical performance of greenhouse extraction in typical areas through the integration of semantic segmentation and edge detection. This method fully utilizes the results of both greenhouse surface and edge detection. By employing the edge constraint method, the system produces vector patches that closely conform to the actual morphology of the greenhouse, achieving refined extraction. Figure 8(a) illustrates a scenario of extreme density in typical greenhouses. If the vectorization is directly applied to the semantic segmentation results, it would be challenging to accurately representing the morphology and quantity of the greenhouses in this area. In the extraction process of this method, the greenhouse-covered areas are first accurately identified by the semantic segmentation model, where contiguous greenhouses are recognized as a whole. Then, with the precise extraction of greenhouse boundaries using the edge detection model, including both "greenhouse–land boundaries" and "greenhouse–greenhouse boundaries," clusters of greenhouses can be clearly distinguished. Subsequently,
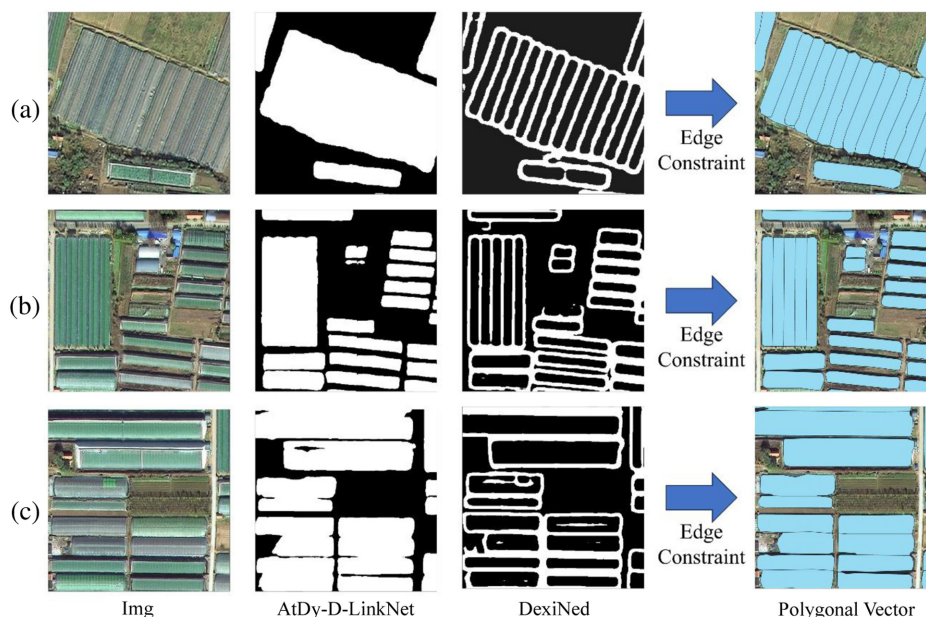

Fig. 8 Effect diagram of edge constraints in AG intensive areas. This figure presents three regions: (a)–(c). Each region showcases semantic segmentation using AtDy-D-LinkNet, edge detection using DexiNed, and polygonal vector output using edge constraint.

Table 3 Quantity extraction accuracy of typical zones.

Typical zones	N_{gt}	Method	N_{pre}	QA
Zone 1	1103	Only semantic segmentation	706	0.6401
		With edge constraint	1084	0.9828
Zone 2	824	Only semantic segmentation	493	0.5983
		With edge constraint	795	0.9648
Zone 3	1124	Only semantic segmentation	998	0.8879
		With edge constraint	1091	0.9706

the edge constraint method proposed in Sec. 3.3 is applied, utilizing the edge characteristics of the greenhouse to constrain the greenhouse surface, ultimately generating refined greenhouse vector patches.

The design intent of the edge constraint method is to compensate for the limited edge perception capability of the semantic segmentation model and to address edge adhesion through model collaboration. In fact, the process of edge constraint not only effectively separates adhering patches of greenhouses but also naturally optimizes and mitigates some deficiencies in the model extraction results. As shown in Fig. 8(b), there is evident underextraction in the segmentation results, while the edge detection model accurately identifies the boundaries of the underextracted greenhouses. The combination of these results ensures their complete preservation in the vector patches. In Fig. 8(c), spectral changes caused by specular reflections on the greenhouse roofs lead to holes in the segmentation results. The edge detection model is highly sensitive to the visual features of the sharply changing greenhouse roofs, which may lead to potential misextraction. These issues are naturally resolved during the vector processing of edge constraint.

The experiments demonstrate that this method can significantly improve the accuracy of greenhouse quantity extraction. When converting the raster results extracted by the model into vectors and counting the number of patches, according to Table 3, in typical areas, such as zone 1 and zone 2, with dense distribution, the number of greenhouses obtained solely through semantic segmentation is much lower than the actual number. However, our method can separate adhering greenhouses, leading to a 34.2% and 36.6% improvement in QA for zone 1 and zone 2, respectively. In zone 3, where greenhouses are relatively scattered and only a few are densely distributed, the semantic segmentation model can accurately identify the dispersed greenhouses, achieving high-QA. On this basis, our method improves by 8.2%.

4.4 Greenhouse Information Statistics and Mapping

Based on the imagery of the study area from October 2021, this method was used to extract information on the area, quantity, and spatial distribution of greenhouses. According to the statistics, the greenhouse coverage area in the study area is ~232.72 km^2. Using the edge constraint method and vector statistics, the quantity of greenhouses extracted is 157,170. The distribution of greenhouses in the Shouguang area is closely related to human activities and geographical location. Supported by government policies, villages and towns with moderate population density and proximity to the city center are ideal locations for greenhouse construction. According to Fig. 9, greenhouses in Shouguang are mainly distributed in the central and southern regions, primarily featuring nondense greenhouse distribution types, with dense greenhouse distribution types being relatively rare. Overall, the results for area extraction and quantity extraction demonstrate strong consistency.

5 Discussion

This study addresses the practical need for extracting the quantity, area, and spatial distribution of AGs. We propose a method for integrating semantic segmentation and edge detection for AG extraction. Through experiments and comparative analysis in the study area, we discuss the following points.
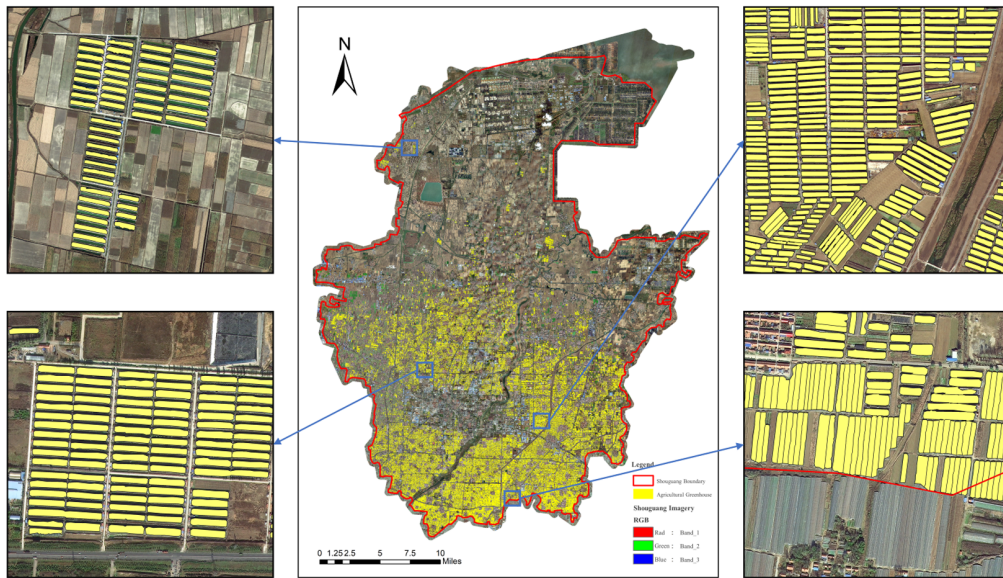
Fig. 9 Spatial distribution of greenhouses in Shouguang City.

(1) Due to the dense spatial distribution of AGs, which often form compact clusters with narrow intervals between them, medium- and low-resolution imagery struggles to accurately delineate greenhouse boundaries, leading to low extraction accuracy. This study utilizes 0.5 m high-resolution remote sensing imagery as the data source. This imagery offers finer pixels and captures richer spatial and textural information, providing a solid data foundation for accurate greenhouse identification.

(2) To achieve precise greenhouse segmentation, we propose the AtDy-D-LinkNet, based on D-LinkNet and enhanced with the CBAM attention module and DySample dynamic sampling strategy. Experimental evaluation results indicate that this model effectively learns and identifies greenhouse features, is adept at filtering out noise interference, and reduces internal holes, false positives, and false negatives. Compared to other models, the extraction results of this model show significant improvements in both objective accuracy metrics and subjective visual assessment.

(3) This study constructs an edge constraint method as a bridge to integrate the results of semantic segmentation and edge detection, effectively combining the advantages of both models. This approach addresses the edge adhesion problem that arises in dense greenhouse clusters predicted by a single semantic segmentation model, achieving high-accuracy extraction of greenhouse quantity and areas. However, compared to traditional methods, this approach requires comprehensive training of both deep learning models, which affects efficiency.

(4) The predicted results of greenhouses are coarse raster data, which cannot be directly used for greenhouse monitoring and management tasks. Previous studies have lacked further geographic processing of the predicted results. This study optimizes and integrates the predicted results into vector data through postprocessing. It aims to obtain vector results that closely match the actual area, quantity, and spatial distribution of greenhouses, thereby facilitating advanced analysis and decision-making.

6 Conclusions

The widespread adoption of AGs has promoted local agricultural and economic development but has also led to environmental pollution issues. Timely and accurate extraction of greenhouse areas, quantity, and spatial distribution information is crucial for the sustainable development of the local greenhouse economy. Traditional semantic segmentation methods tend to produce edge adhesion in greenhouses, making it difficult to accurately extract their numbers. This paper proposes a more effective semantic segmentation model, AtDy-D-LinkNet, and develops a method for AG extraction that integrates semantic segmentation and edge detection. We verified

the scientific validity and feasibility of this method using submeter high-resolution remote sensing imagery in Shouguang City, Shandong Province. The experimental results indicate that: (1) The AtDy-D-LinkNet model demonstrates high extraction accuracy, surpassing traditional models. (2) Edge constraint method effectively addresses the edge adhesion problem caused by segmentation, enabling precise extraction of greenhouse areas, numbers, and spatial distribution. The final extraction results show that as of October 2021, the greenhouse coverage area in Shouguang City is ∼232.72 km^2, with about 157,170 greenhouses. Spatially, the greenhouses are primarily distributed in the central and southern regions of Shouguang City.

The precise extraction of greenhouses based on deep learning requires a significant number of samples, which necessitates extensive and labor-intensive manual annotation. Therefore, future research will focus on developing greenhouse extraction methods suitable for small sample scenarios to advance the automation of remote sensing interpretation.

Code and Data Availability

The data that support the findings of this article are not publicly available due to commercial restrictions. They can be requested from the author at e-mail: z22160031@s.upc.edu.cn

Acknowledgments

References

1. Y. Liu and Y. Zhou, "Reflections on China's food security and land use policy under rapid urbanization," *Land Use Policy* **109**, 105699 (2021).
2. Y. H. Ma et al., "The situation of grain production in China and the protection of cultivated land," *Macroecon. Manage.* **2023**(9), 61–70 (2023).
3. W. F. Fang and Z. X. Kong, "Investigation and analysis report on capital investment and economic benefits of vegetables-taking Shouguang greenhouse vegetables as an example," *World Surv. Res.* **2012**(6), 21–24 (2012).
4. H. C. J. Godfray et al., "Food security: the challenge of feeding 9 billion people," *Science* **327**(5967), 812–818 (2010).
5. Y. P. Chen et al., "Analysis of present situation and control of heavy metal pollution in vegetable greenhouse soils," *J. Agro-Environ. Sci.* **37**(1), 9–17 (2018).
6. L. Yang et al., "Assessment and source identification of trace metals in the soils of greenhouse vegetable production in eastern China," *Ecotoxicol. Environ. Saf.* **97**, 204–209 (2013).
7. S. Parra, F. J. Aguilar, and J. Calatrava, "Decision modelling for environmental protection: the contingent valuation method applied to greenhouse waste management," *Biosyst. Eng.* **99**(4), 469–477 (2008).
8. P. Picuno, "Innovative material and improved technical design for a sustainable exploitation of agricultural plastic film," *Polym. Plast. Technol. Eng.*, **53**(10), 1000–1011 (2014).
9. Z. Steinmetz et al., "Plastic mulching in agriculture. Trading short-term agronomic benefits for long-term soil degradation?" *Sci. Total Environ.* **550**, 690–705 (2016).
10. S. Pearson, A. E. Wheldon, and P. Hadley, "Radiation transmission and fluorescence of nine greenhouse cladding materials," *J. Agric. Eng. Res.* **62**(1), 61–69 (1995).
11. D. Koc-San, "Evaluation of different classification techniques for the detection of glass and plastic greenhouses from WorldView-2 satellite imagery," *J. Appl. Remote Sens.* **7**(1), 073553 (2013).
12. L. Lu, L. Di, and Y. Ye, "A decision-tree classifier for extracting transparent plastic-mulched landcover from Landsat-5 TM images," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **7**(11), 4548–4558 (2014).
13. H. Sun et al., "Mapping plastic greenhouses with two-temporal Sentinel-2 images and 1D-CNN deep learning," *Remote Sens.* **13**(14), 2820 (2021).
14. J. Wu et al., "Plastic greenhouse recognition based on GF-2 data and multi-texture features," *Trans. Chin. Soc. Agric. Eng.* **35**, 173–183 (2019).
15. L. Ma et al., "Deep learning in remote sensing applications: a meta-analysis and review," *ISPRS J. Photogramm. Remote Sens.* **152**, 166–177 (2019).
16. T. Fu et al., "Using convolutional neural network to identify irregular segmentation objects from very high-resolution remote sensing imagery," *J. Appl. Remote Sens.* **12**(2), 025010 (2018).
17. H. Su et al., "Using improved DeepLabv3+ network integrated with normalized difference water index to extract water bodies in Sentinel-2A urban remote sensing images," *J. Appl. Remote Sens.* **15**(1), 018504 (2021).

18. M. Kaur, R. Singh, and H. Gritli, "Special section guest editorial: Meeting the challenges of ecosystem management using remote sensing," *J. Appl. Remote Sens.* **17**(2), 022201 (2023).

19. S. Khanal et al., "Remote sensing in agriculture—accomplishments, limitations, and opportunities," *Remote Sens.* **12**(22), 3783 (2020).

20. Q. Yang et al., "Mapping plastic mulched farmland for high resolution images of unmanned aerial vehicle using deep semantic segmentation," *Remote Sens.* **11**(17), 2008 (2019).

21. Z. Chen et al., "A convolutional neural network for large-scale greenhouse extraction from satellite images considering spatial features," *Remote Sens.* **14**(19), 4908 (2022).

22. X. L. Zhang et al., "UAV images agricultural greenhouse information extraction method based on DE-Segformer," *Eng. Surv. Mapp.* **33**(2), 56–64 (2024).

23. H. Li et al., "EAGNet: a method for automatic extraction of agricultural greenhouses from high spatial resolution remote sensing images based on hybrid multi-attention," *Comput. Electron. Agric.* **202**, 107431 (2022).

24. X. Zhang et al., "High-resolution boundary refined convolutional neural network for automatic agricultural greenhouses extraction from GaoFen-2 satellite imageries," *Remote Sens.* **13**(21), 4237 (2021).

25. J. Feng et al., "PODD: a dual-task detection for greenhouse extraction based on deep learning," *Remote Sens.* **14**(19), 5064 (2022).

26. K. Sun et al., "Deep high-resolution representation learning for human pose estimation," in *IEEE/CVF Conf. Comput. Vision and Pattern Recognit. (CVPR)*, Long Beach, California, pp. 5686–5696 (2019).

27. Q. Wang et al., "Simultaneous extracting area and quantity of agricultural greenhouses in large scale with deep learning method and high-resolution remote sensing images," *Sci. Total Environ.* **872**, 162229 (2023).

28. S. Woo et al., "CBAM: convolutional block attention module," in *Proc. Eur. Conf. Comput. Vision (ECCV)*, pp. 3–19 (2018).

29. W. Liu et al., "Learning to upsample by learning to sample," in *Proc. IEEE/CVF Int. Conf. Comput. Vision*, pp. 6027–6037 (2023).

30. N. Zioulis et al., "Hybrid skip: a biologically inspired skip connection for the UNet architecture," *IEEE Access* **10**, 53928–53939 (2022).

31. J. Kittler, "On the accuracy of the Sobel edge detector," *Image Vision Comput.* **1**(1), 37–42 (1983).

32. J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-8**(6), 679–698 (1986).

33. P. Dollar and C. L. Zitnick, "Fast edge detection using structured forests," *IEEE Trans. Pattern Anal. Mach. Intell.* **37**(8), 1558–1570 (2015).

34. P. Arbeláez et al., "Contour detection and hierarchical image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(5), 898–916 (2011).

35. S. Xie and Z. Tu, "Holistically-nested edge detection," in *IEEE Int. Conf. Comput. Vision (ICCV)*, Santiago, Chile, pp. 1395–1403 (2015).

36. X. S. Poma, E. Riba, and A. Sappa, "Dense extreme inception network: towards a robust CNN model for edge detection," in *Proc. IEEE/CVF Winter Conf. Appl. of Comput. Vision*, pp. 1923–1932 (2020).

37. D. Chollet, "Xception: deep learning with depthwise separable convolutions," in *IEEE Conf. Comput. Vision and Pattern Recognit. (CVPR)*, *Honolulu, Hawaii*, IEEE, pp. 1800–1807 (2017).

Yawen He is an associate professor at China University of Petroleum (East China), specializing in intelligent interpretation of agricultural remote sensing and marine geographic information systems.

Feng Jin is a graduate student at China University of Petroleum (East China) and is interested in artificial intelligence technologies and primarily researches intelligent remote sensing interpretation and agricultural remote sensing.

Yongheng Li is a graduate student at China University of Petroleum (East China), focusing on marine geographic information systems and marine big data mining.