

## Article

# Artemisia Frigida Distribution Mapping in Grassland with Unmanned Aerial Vehicle Imagery and Deep Learning

Yongcai Wang<sup>1</sup>, Huawei Wan<sup>2,\*</sup>, Zhuowei Hu<sup>1</sup>, Jixi Gao<sup>2</sup>, Chenxi Sun<sup>2</sup> and Bin Yang<sup>3</sup>

<sup>1</sup> College of Resource Environment and Tourism, Capital Normal University, Beijing 100048, China; 2210901019@cnu.edu.cn (Y.W.); huzhuowei@cnu.edu.cn (Z.H.)

<sup>2</sup> Satellite Application Center for Ecology and Environment, Ministry of Ecology and Environment of the People's Republic of China, Beijing 100094, China; gjx@nies.org (J.G.); sunchenxi@secmep.cn (C.S.)

<sup>3</sup> Ysense Information Technology and Equipment Inc., Qingdao 266000, China; yangbin@yusense.com.cn

\* Correspondence: wanhw@secmep.cn

**Abstract:** *Artemisia frigida*, as an important indicator species of grassland degradation, holds significant guidance significance for understanding grassland degradation status and conducting grassland restoration. Therefore, conducting rapid surveys and monitoring it is crucial. In this study, to address the issue of insufficient identification accuracy due to the large density and small size of *Artemisia frigida* in UAV images, we improved the YOLOv7 object detection algorithm to enhance the performance of the YOLOv7 model in *Artemisia frigida* detection. We applied the improved model to the detection of *Artemisia frigida* across the entire experimental area, achieving spatial mapping of *Artemisia frigida* distribution. The results indicate: In comparison across different models, the improved YOLOv7 + Biformer + wise-iou model exhibited the most notable enhancement in precision metrics compared to the original YOLOv7, showing a 6% increase. The mean average precision at intersection over union (IoU) threshold of 0.5 (mAP@.5) also increased by 3%. In terms of inference speed, it ranked second among the four models, only trailing behind YOLOv7 + biformer. The YOLOv7 + biformer + wise-iou model achieved an overall detection precision of 96% and a recall of 94% across 10 plots. The model demonstrated superior overall detection performance. The enhanced YOLOv7 exhibited superior performance in *Artemisia frigida* detection, meeting the need for rapid mapping of *Artemisia frigida* distribution based on UAV images. This improvement is expected to contribute to enhancing the efficiency of UAV-based surveys and monitoring of grassland degradation. These findings emphasize the effectiveness of the improved YOLOv7 + Biformer + wise-iou model in enhancing precision metrics, overall detection performance, and its applicability to efficiently map the distribution of *Artemisia frigida* in UAV imagery for grassland degradation surveys and monitoring.

**Keywords:** *Artemisia frigida*; unmanned aerial vehicle; object detection; deep learning; density distribution



**Citation:** Wang, Y.; Wan, H.; Hu, Z.; Gao, J.; Sun, C.; Yang, B. *Artemisia Frigida* Distribution Mapping in Grassland with Unmanned Aerial Vehicle Imagery and Deep Learning. *Drones* **2024**, *8*, 151. <https://doi.org/10.3390/drones8040151>

Academic Editor: Yangquan Chen

Received: 23 January 2024

Revised: 27 March 2024

Accepted: 31 March 2024

Published: 15 April 2024



**Copyright:** © 2024 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

Indicator species of grassland degradation play a crucial role in community succession and grassland degradation warning within grassland ecosystems. They are of significant importance in studying the response of grasslands to grazing, human disturbances, and other factors leading to degradation [1]. As indicators of grassland degradation, monitoring changes in various indicators of these indicator species, such as importance value and dominance, has become essential in assessing degraded grasslands [2]. These key parameters of grassland degradation indicator species can be utilized to evaluate the health and ecological functionality of grasslands [3].

Given the role of indicator species in the study of grassland ecosystem degradation, research has identified various grassland degradation indicator species, such as *Iris lactea* var. *chinensis*, *Stipa breviflora*, *Convolvulus ammannii*, *Artemisia frigida*, and others. *Artemisia frigida* is a small semi-shrub belonging to the Asteraceae family. It is a facultative clonal

plant, resilient to drought, trampling, and soil erosion, with strong regenerative capabilities through rooting and sprouting [4]. As one of the dominant species in desert grasslands, *Artemisia frigida* significantly influences the structure and function of the community. Serving as an important indicator species for grassland degradation, the dynamic changes in *Artemisia frigida* to some extent reflect grassland degradation [5]. Therefore, conducting surveys and monitoring of *Artemisia frigida* contributes to understanding the status of grassland degradation. Currently, investigations of degradation indicator species like *Artemisia frigida* heavily rely on manual field surveys and records. However, these surveys may be prone to noise due to the influence of staff knowledge and working conditions. Additionally, manual surveys are time-consuming and labor-intensive, making large-scale rapid surveys challenging [6].

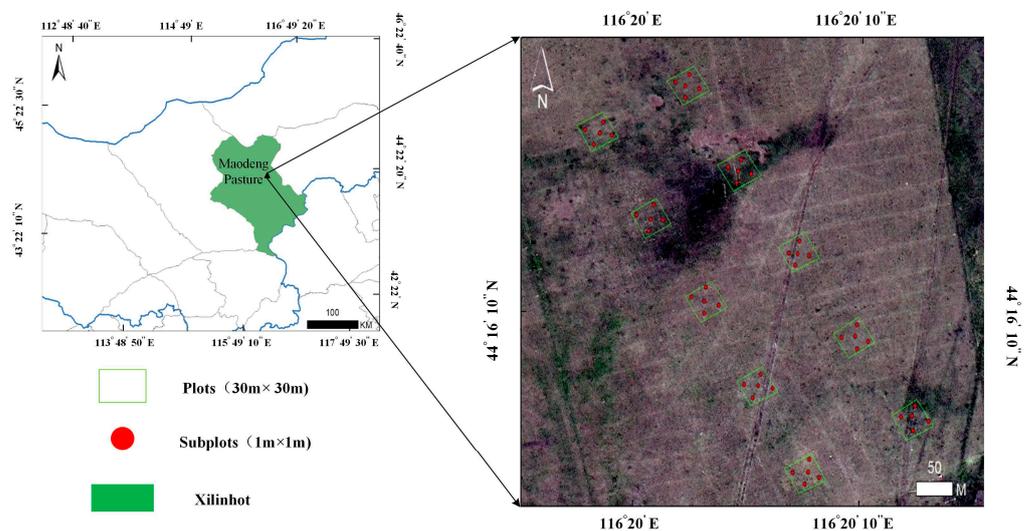
In addition to field surveys, there is also relevant research utilizing remote sensing technology for the classification and identification of degradation species such as *Artemisia frigida*. Currently, research primarily focuses on fine plant classification through ground spectral measurements within a small area. Conducting spectral measurements on a larger scale still faces several challenges. For example, Gai et al. [7] used canopy spectral information to identify plants such as *Artemisia frigida* and *Lilium brownii*. Niu Yalong et al. [8] measured the spectra of five typical sand vegetation types, including *Artemisia frigida*, and analyzed the similarities and differences between them. Unmanned aerial vehicle (UAV) remote sensing offers maneuverability and flexibility, and through carrying multiple sensors, it enables the acquisition of multi-source data. It can provide centimeter-level ultra-high spatial resolution data [9], which is suitable for fine plant classification and identification. Therefore, UAV remote sensing has found wide applications in grassland plant classification and ecological parameter acquisition [10,11]. For instance, Yang et al. [12] obtained images using a UAV hyperspectral imaging system. They constructed classification features for species in desert grasslands through spectral transformations of vegetation indices. Utilizing a decision tree model, they classified and identified key species such as *Stipa breviflora*, *Artemisia frigida*, and *Salsola collina* in desert grassland. The above-mentioned research on grassland plant identification and classification using UAVs mainly focuses on the scale of grassland communities, with few studies conducted at the level of plant populations or individuals for recognition.

The high-resolution images obtained by unmanned aerial vehicles (UAVs) provide excellent data support for grassland plant recognition and identification. However, for small and dense objects like *Artemisia frigida*, conventional image processing algorithms have limited recognition capabilities, resulting in relatively low identification accuracy [13]. With the development of deep learning technology, numerous object detection algorithms have been developed. These detection methods, trained on massive datasets and tested on small sample sets, extract deep features from data samples, demonstrating strong learning capabilities and high accuracy in object detection and recognition [14]. Among these object detection algorithms, YOLO (You Only Look Once) represents a single-stage object detection algorithm [15]. Compared to two-stage object detection algorithms, its major advantages lie in fast execution speed and high detection accuracy. However, for grassland plants with dense distribution and relatively small size compared to trees or potted plants, detection algorithms face higher demands in recognizing densely packed small targets. Considering this, this study, which is based on UAV images, improves the detection performance of the YOLOv7 model for densely packed small objects. The enhanced model is then applied to the detection of *Artemisia frigida* in sample plots. Spatial distribution and density mapping and analysis of *Artemisia frigida* in sample plots are conducted, aiming to provide technical references for the investigation of degradation indicator species such as *Artemisia frigida*.

## 2. Materials and Methods

### 2.1. Study Area

The Xilingol Grassland is located in the eastern part of the Xilingol Plateau in Inner Mongolia Autonomous Region, China. It is one of the largest grasslands in northern China, characterized by a predominantly continental arid and semi-arid climate. The elevation ranges from 760 to 1926 m, with an average annual temperature of 0 to 3 °C and an annual precipitation of 150 to 350 mm. The precipitation is concentrated mainly from July to September, exhibiting distinct seasonal characteristics with warm summers and cold winters. The grassland covers an area of approximately 179,600 square kilometers and includes diverse grassland types such as meadow grassland, typical grassland, and dune desert grassland. The experimental area selected for this study is located in the Maodeng Ranch, Xilinhot City, Inner Mongolia Autonomous Region, China (116.03° to 116.50° E, 44.80° to 44.82° N), as shown in Figure 1. It is a typical grassland and is renowned as one of the prominent ranches on the Xilinhot Grassland. This area features a rich variety of plant species, with dominant vegetation including Poaceae, Fabaceae, and Asteraceae pasture grasses.



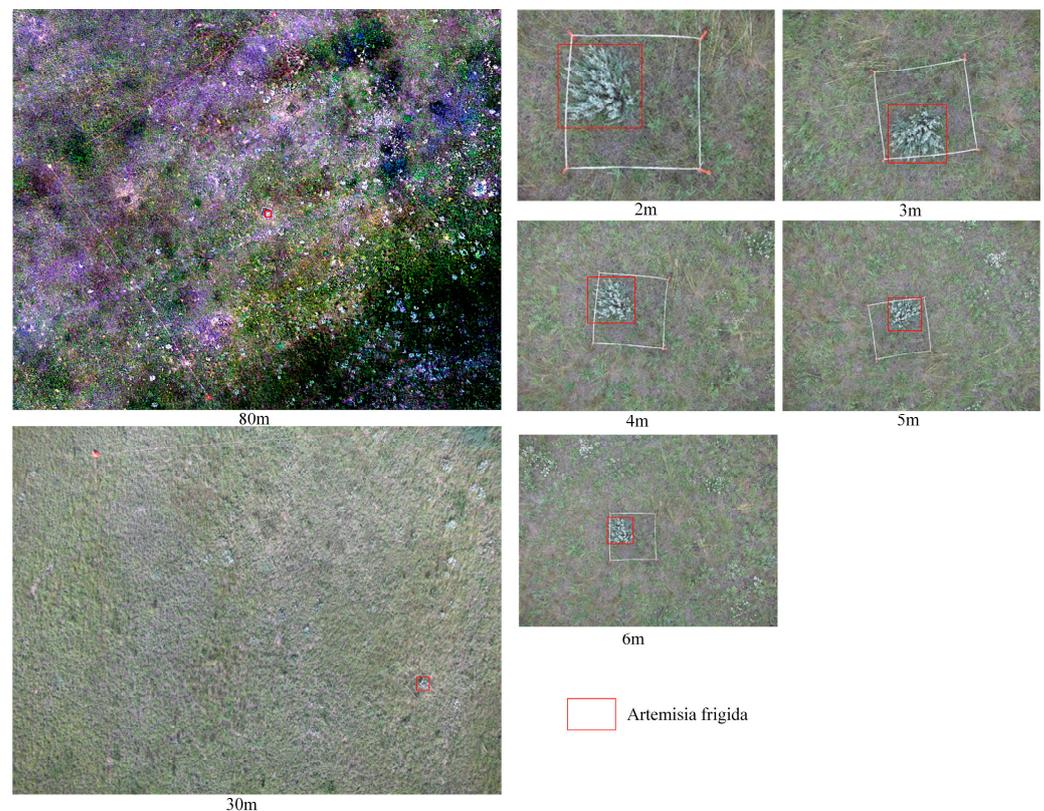
**Figure 1.** Study area.

### 2.2. Data

We collected data in August 2022 using the DJI M300 RTK UAV (DJI, Inc., Shenzhen, China). The UAV adopts a new flight control system and motor system, with the characteristics of high-precision positioning, high stability and reliability. It has an endurance of approximately 55 min, can withstand a maximum wind speed of 15 m/s, has a maximum horizontal flight speed of 23 m/s, and a maximum range of 15 km. Equipped with the AQ600 multispectral camera (Yusense, Inc., Qingdao, China), this camera is lightweight, easy to install, and easy to operate, compatible with various UAV. It consists of one 12.3-megapixel RGB channel and five 3.2-megapixel multispectral channels. The RGB channel has a focal length of 7.2 mm, a field of view of  $47.4^\circ \times 36.4^\circ$ , and the imaging resolution is  $4056 \times 3040$  pixels. The ground sampling distance for the RGB channel is 1.76 cm @ h80m. The five spectral bands of the multispectral channels are as follows: blue ( $450 \text{ nm} \pm 35 \text{ nm}$ ), green ( $555 \text{ nm} \pm 27 \text{ nm}$ ), red ( $660 \text{ nm} \pm 22 \text{ nm}$ ), red edge ( $720 \text{ nm} \pm 10 \text{ nm}$ ), and near-infrared ( $840 \text{ nm} \pm 30 \text{ nm}$ ), and the imaging resolution is  $2048 \times 1536$  pixels. The ground sampling distance for the multispectral channel is 2.52 cm @ h80m. Given the ability to recognize target species, to improve efficiency in field data collection, we set the flight altitude to 80 m, flight speed to 3 m/s, and the lateral and longitudinal overlap rates to 85%. A total of 23 flight lines are set, covering an area of  $500 \times 500$  m. At a height of 80 m, the ground resolution of multispectral channel imagery

is 2.52 cm, making it difficult to capture the smaller *Artemisia frigida*. Therefore, in this study, we only used RGB channel imagery for *Artemisia frigida* detection.

Before the flight, we set up 10 plots of  $30 \times 30$  m each within a  $500 \times 500$  m area. Within each plot, we placed 5 sample quadrats of  $1 \times 1$  m each and recorded the coordinates of each plot and subplot. The distribution of plots and quadrats can be seen in Figure 1. We conducted species surveys in each plot, meticulously recording the number of *Artemisia frigida* within them. This data was used to validate the accuracy of *Artemisia frigida* detection at the plot scale in subsequent analyses. To facilitate annotation in the future, we captured images of each plot at a height of 30 m, which serve as reference images for image annotation. Additionally, within each  $30 \times 30$  m plot, we set 5 subplots and recorded the coordinates of each subplot. We then captured images within and around these subplots at heights ranging from 2–6 m with 1 m intervals, as shown in Figure 2. These images serve as references for annotation.



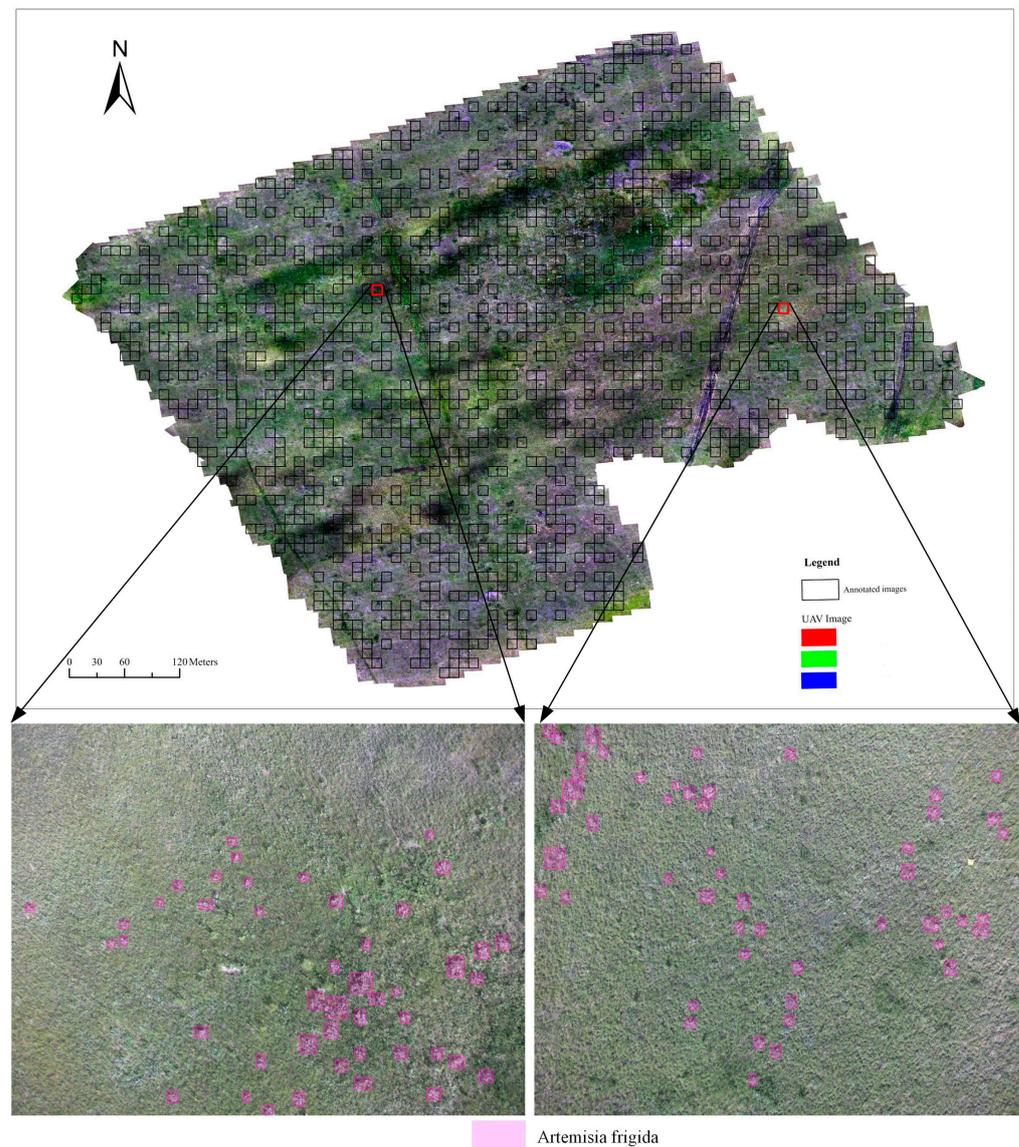
**Figure 2.** *Artemisia frigida* images at different heights.

### 2.3. Data Preprocessing and Annotation

A total of 790 RGB images were obtained during the flight. After the flight, the relative positions and orientations of the aerial images were reconstructed and merged into a large orthomosaic. An orthomosaic is a visual representation of an area, created from many photos that were stitched together in a geometrically corrected manner. We utilized the Structure from Motion approach implemented in Agisoft Metashape Professional version 2.0.2 for this purpose. Agisoft Metashape processes all aerial images as input and aligns them through bundle adjustment, enabling the generation of a point cloud representing the topography of the surveyed area. From this point cloud, a digital surface model was created to orthorectify the orthomosaic. During orthomosaic generation, we disabled blending to preserve the original image information without smearing. The final orthomosaic we generated was exported with a ground sample distance of 1 cm.

To ensure the quality of data annotation, we established an *Artemisia frigida* UAV image annotation reference library based on ground plot survey data and UAV images captured at

different heights. We randomly cropped 790 images captured at 80 m height, with a size of  $4056 \times 3040$  pixels, resulting in a total of 4150 images with a size of  $1280 \times 1280$  pixels. Out of these, 1200 images were randomly selected for annotation, as shown in Figure 3, which was completed in the Label Studio annotation platform. During annotation, the *Artemisia frigida* UAV image annotation library was referenced, and comparison was made with annotated images falling within the survey plot range and field survey *Artemisia frigida*, ensuring the quality of annotation. The annotations were exported in YOLO format. The annotated images were divided into training, validation, and test sets, with 1000 images in the training set, 100 images in the validation set, and 100 images in the test set. Additionally, to expand the training set, we performed augmentation on the training images. The augmentation techniques included horizontal flipping, vertical flipping, random cropping, random translation scaling rotation, center cropping, and elastic transformation.



**Figure 3.** Annotation example.

#### 2.4. YOLOv7 Improvement

The YOLO series models utilize an end-to-end approach for object detection and localization in images, predicting both categories and positions simultaneously. A typical YOLO object detector comprises four main components: input, backbone network, neck, and head. To enhance training data diversity, data augmentation techniques are applied in

the input module during the training phase. The diversified data then undergoes feature extraction at different scales in the backbone module. The neck module incorporates upsampling and feature concatenation layers for feature injection, providing additional details for the final stage. The head module compares the predicted object categories and positions with ground truth labels to generate the loss function results. Subsequently, backpropagation updates parameters in the backbone, neck, and head modules. Model training continues based on input data until the loss function stabilizes, indicating the completion of the training process [16–18]. In terms of architecture, the YOLOv7 model introduces an extended efficient layer aggregation network ELAN-based extension known as E-ELAN. E-ELAN utilizes Expand, Shuffle, and Merge Cardinality to continuously enhance the network’s learning capability without destroying existing gradient paths. In terms of architecture, E-ELAN only changes the architecture of the computation blocks, while the architecture of the transition layer remains unchanged. Our strategy is to use group convolutions to expand the channels and cardinality of the computation blocks. We apply the same group parameters and channel multipliers to all computation blocks in the computation layer. Then, based on the group parameter  $g$ , we shuffle the feature maps computed by each computation block into  $g$  groups and then concatenate them together. At this point, the number of channels in each group of feature maps will be the same as the number of channels in the original architecture. Finally, we add  $g$  groups of feature maps to perform merge cardinality. In addition to maintaining the original ELAN design architecture, E-ELAN can also guide computation modules from different groups to learn more diversified features [19]. Despite being considered one of the top-tier object detection models, YOLOv7 still faces challenges in detecting dense small objects. To address this, the BiFormer module is introduced into the YOLOv7 backbone network to improve the detection capability of dense small objects. Simultaneously, to mitigate the impact of low-quality samples and enhance the overall detection performance of the model, the box loss of the YOLOv7 model is replaced with Wise-IoU (WIoU). The improved YOLOv7 network architecture is illustrated in Figure 4.

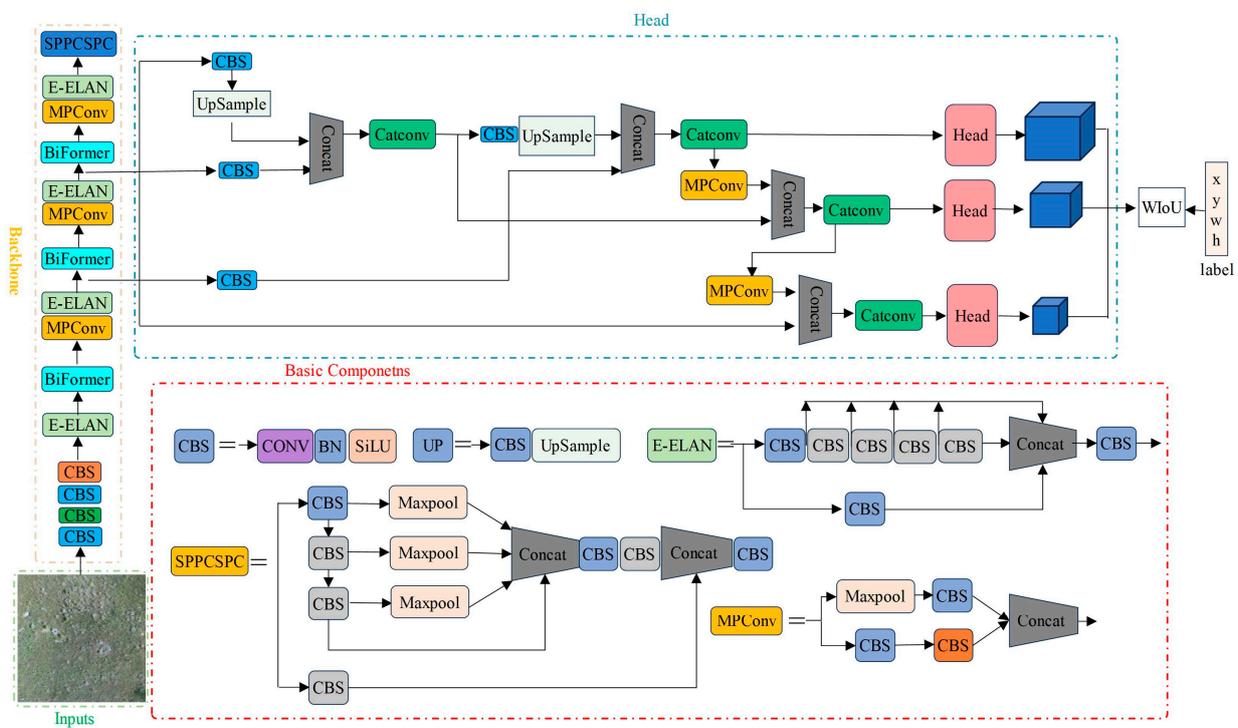


Figure 4. The improved YOLOv7 network architecture.

The network architecture of BiFormer consists of the Bi-Level Routing Attention (BRA) module, as shown in Figure 5. Due to its adaptive attention mechanism, which focuses on a small portion of relevant tokens without dispersing attention to unrelated tokens, it exhibits excellent performance and high computational efficiency in dense prediction tasks. This architecture follows the design principles of most vision transformer architectures and adopts a four-level pyramid structure, namely, a  $32\times$  downsampling. In the first stage, BiFormer employs overlapping block embeddings, while in the second to fourth stages, block merging modules are used to reduce the input spatial resolution while increasing the number of channels. Subsequently, continuous BiFormer blocks are applied for feature transformation. It is important to note that at the beginning of each block, depthwise convolutions are used to implicitly encode relative positional information. Following this, BRA modules and Multi-Layer Perceptron (MLP) modules with an expansion rate of 2 are sequentially applied, and used for modeling cross-location relationships and embedding at each position, respectively [20,21].

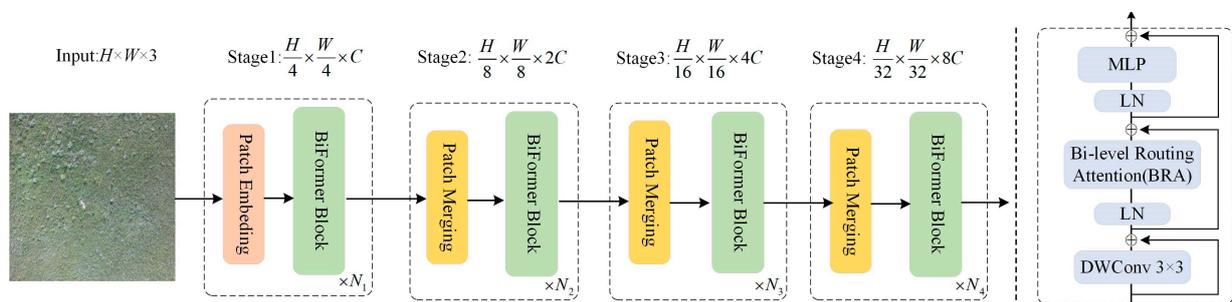


Figure 5. The structure of BiFormer.

The BRA module is a different dynamic sparse attention mechanism that enables more flexible computation allocation and content awareness [20], allowing the model to have dynamic query-aware sparsity. The BRA module is shown in Figure 6. The process of BRA can be divided into three steps. Firstly, assuming we input a feature map, it is divided into multiple regions, and queries, keys, and values are obtained through linear projection. Secondly, we use an adjacency matrix to construct a directed graph to find the participating relationships corresponding to different key-value pairs, which can be understood as the regions each given region should participate in. Lastly, with the region-to-region routing index matrix, fine-grained token-to-token attention can be applied.

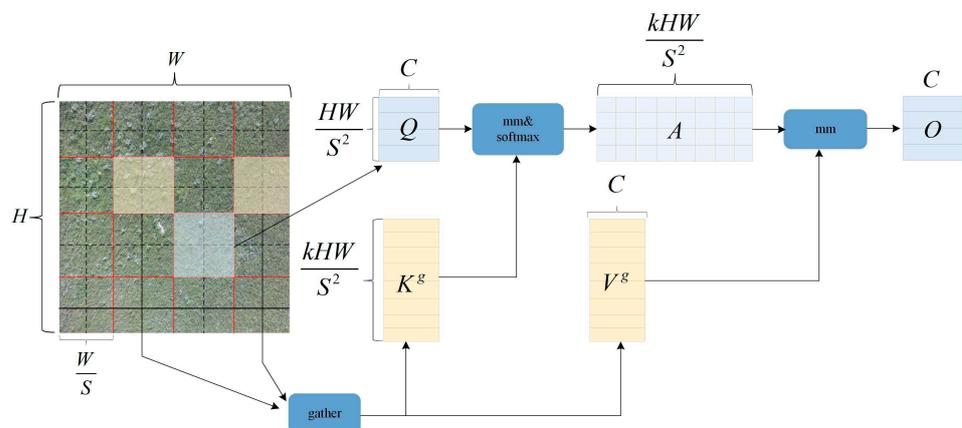


Figure 6. BRA module.

The bounding box loss function, a critical component of object detection loss, plays a pivotal role in significantly enhancing the performance of object detection models when well-defined. Existing research has predominantly operated under the assumption of high-quality training data samples, concentrating on improving the fitting capability of

the bounding box loss. However, it has been observed that object detection training sets often include low-quality samples. Blindly intensifying bounding box regression on such low-quality examples poses a risk to localization performance. To address this issue, Focal-EIoU v1 was introduced as a solution [22], but due to its static focusing mechanism (FM), the potential of non-monotonic FM was not fully exploited. Based on this idea, Tong et al. proposed Wise-IoU, incorporating a dynamic non-monotonic FM [23]. The dynamic non-monotonic FM assesses the quality of anchor boxes using the outlier degree instead of intersection over union (IoU), offering a wise gradient gain allocation strategy. This strategy diminishes the competitiveness of high-quality anchor boxes while mitigating harmful gradients generated by low-quality examples. Consequently, WIoU focuses on anchor boxes of ordinary quality, enhancing the overall performance of the detector.

### 2.5. Model Application and *Artemisia Frigida* Distribution Mapping

Counting plants in representative plots within the experimental area and extrapolating these counts to the entire experimental area is a common method in field plant surveys. However, this sampling survey places high demands on the representativeness of the sampled plots. If the selected survey plots are not representative, it may lead to bias. Conducting global object detection in the experimental area based on high-resolution drone images can to some extent reduce the bias caused by sampling.

To obtain the distribution of *Artemisia frigida* in the experimental area, we divided the orthomosaic of the experimental area into images of size  $1280 \times 1280$  pixels, resulting in a total of 6624 images. The image blocks after slicing are shown in Figure 7. We applied our improved YOLOv7 model to each image to obtain detection bounding boxes for *Artemisia frigida* in each image. The bounding boxes were then transformed to obtain the coordinates of the centroid, which were recorded as the locations of *Artemisia frigida*. To obtain the density of *Artemisia frigida* distribution in the experimental area, we generated a grid of  $10 \times 10$  m size across the entire experimental area and counted the *Artemisia frigida* points in each grid, ultimately obtaining the distribution density of *Artemisia frigida* in the experimental area.

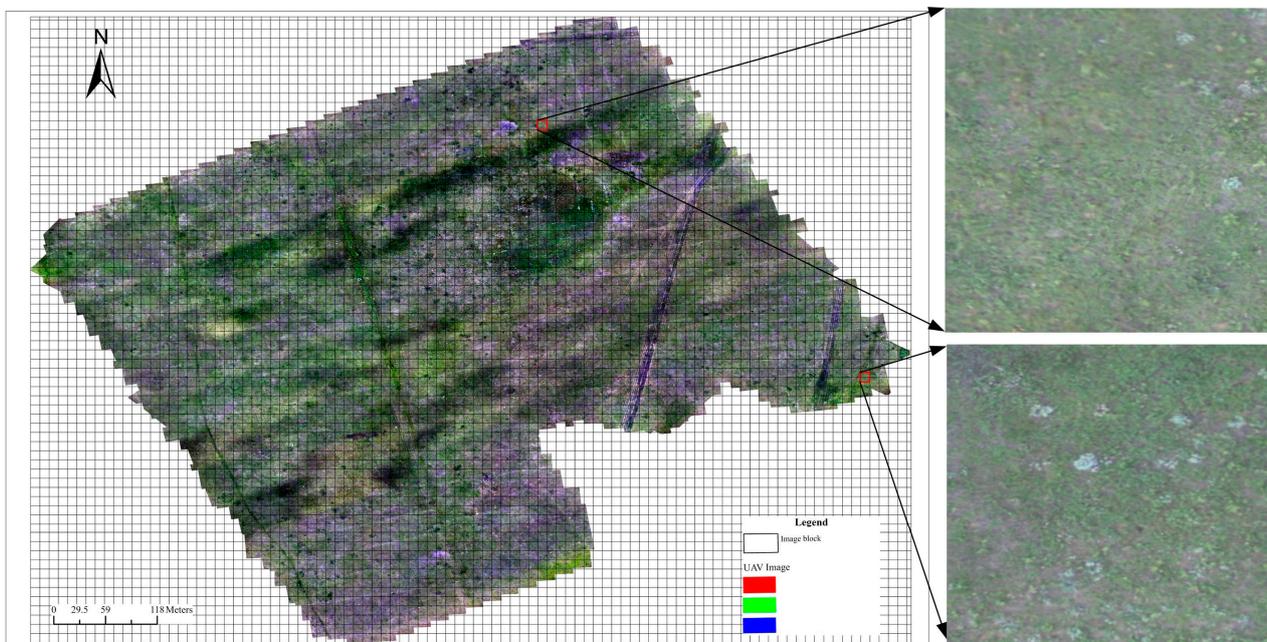


Figure 7. Image blocks after orthophoto slicing.

### 2.6. Experimental Environment and Parameter Settings

The experimental platform for this study is a graphical workstation with an Intel(R) Core(TM) i9-13900K processor operating at 3.00 GHz, 128 GB of RAM, the GPU with the NVIDIA GeForce RTX 4090, 24 GB of VRAM, and the Ubuntu 20.04 operating system. The programs are written in Python 3.8, and model training is conducted on the PyTorch-2.0.0-gpu deep learning framework. The configuration of the experiment plays a crucial role in training deep learning models. The configuration of the experiment for this study is outlined in the Table 1.

**Table 1.** Configuration.

Category	Name	Configuration/Parameter Values
Hardware	CPU	Intel(R) Core(TM) i9-13900K
	GPU	NVIDIA GeForce RTX 4090 24 GB
	Memory	128 GB
Software	CUDA	11.8
	Python	3.8
	PyTorch	2.0
Hyperparameters	Image Size	1280 × 1280
	Learning Rate	0.01
	Learning Rate Decay Frequency	0.1
	Batch Size	6
	Workers	16
	Maximum Training Epochs	300

### 2.7. Evaluation

To evaluate the performance of the model in *Artemisia frigida* detection, five metrics are employed in evaluation. These metrics include precision ( $P$ ), recall ( $R$ ), mean average precision (mAP), parameters, frames per second (FPS). mAP is a metric that better reflects the overall performance of the network. This study utilizes FPS to evaluate the real-time detection performance of the model. A higher FPS indicates faster model detection speed. The calculations for  $P$ ,  $R$ , and mAP are as follows:

$$P = \frac{TP}{TP + FP} \quad (1)$$

$$R = \frac{TP}{TP + FN} \quad (2)$$

$$AP = \int_0^1 P(r) dr \quad (3)$$

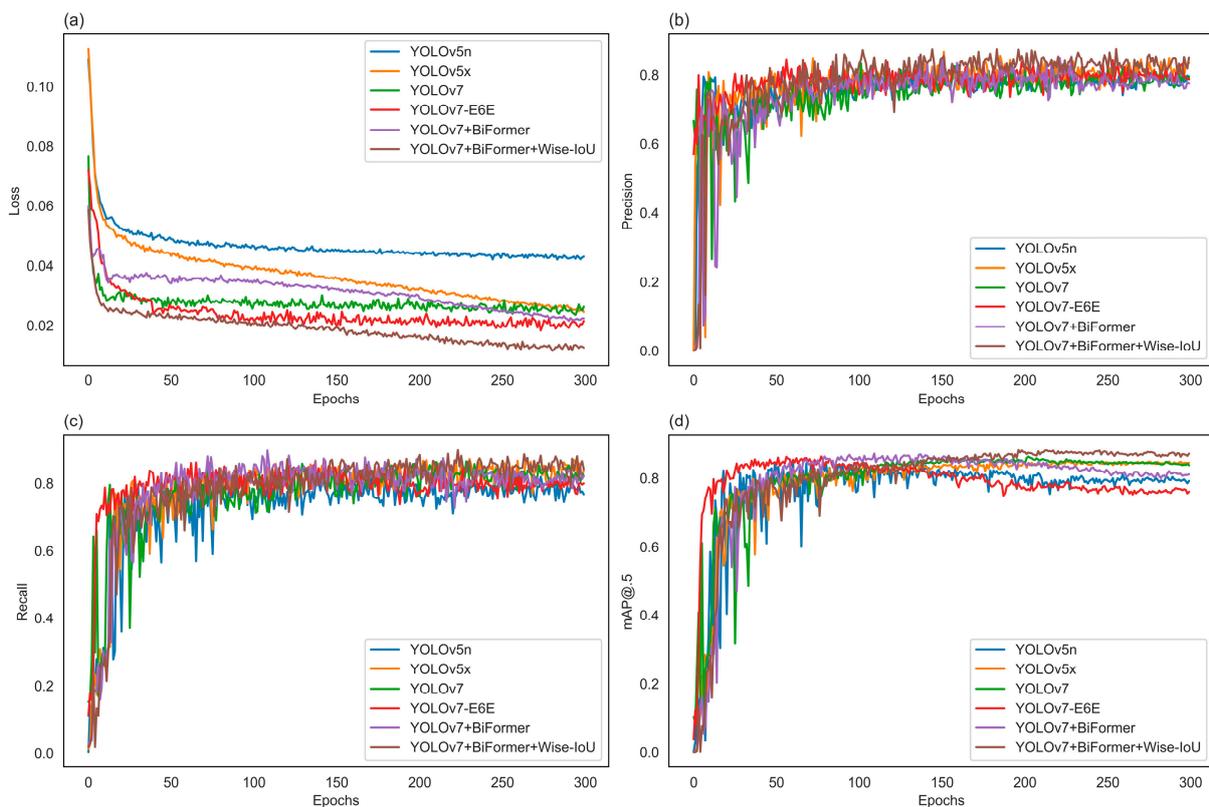
$$mAP = \frac{\sum_1^Q AP(k)}{Q} \quad (4)$$

where True Positive ( $TP$ ) represents accurate predictions made by the model, False Positive ( $FP$ ) indicates incorrect predictions, and False Negative ( $FN$ ) represents instances the model failed to detect.  $AP$  denotes the area enclosed by the precision and recall curve. mAP is the mean of the  $AP$  values for various classes. mAP@.5 refers to the mean average precision (mAP) calculated with an IoU threshold set to 0.5. This means that a detection is considered valid only if the IoU between the predicted bounding box and the ground truth bounding box exceeds 0.5.  $Q$  represents the total number of classes;  $AP(k)$  denotes the Average Precision ( $AP$ ) value for the  $k$ -th class.

### 3. Results

#### 3.1. Model Training and Performance Evaluation

From the change in the model's box loss, as shown in Figure 8, compared to the YOLOv7 + Biforner model, the loss exhibits noticeable oscillations with a gradual descent after a significant spike, indicating a relatively slow convergence speed. In contrast, the loss curve for YOLOv7 + Biforner + Wise-IoU converges more rapidly, improving the model's training speed while achieving a lower loss rate. Examining the changes in model accuracy and recall, the YOLOv7-E6E model shows the fastest initial rise, but after 150 epochs, it experiences a certain degree of decline, possibly due to some low-quality samples. On the other hand, both precision and recall for YOLOv7 + Biforner and YOLOv7 + Biforner + Wise-IoU models steadily increase with training batches.



**Figure 8.** Model evaluation indicators. (a) Box loss; (b) Precision; (c) Recall; (d) mAP@.5.

We conducted a comprehensive evaluation of the improved YOLOv7 model, and the results indicate that the improved model outperforms YOLOv7 in several evaluation metrics. Model performance comparison is shown in Table 2. YOLOv7 + Biforner + Wise-IoU model, compared to the original YOLOv7, the improvement is most notable in the precision metric, with a 6% increase, and the mAP@.5 has increased by 3%. YOLOv7 + Biforner, aside from the precision metric, has also shown a certain degree of improvement compared to YOLOv7, similar to YOLOv7 + Biforner + Wise-IoU. Considering model size and inference speed, the YOLOv7-E6E model has the most parameters and the slowest inference speed, only 12.20 FPS. YOLOv7 + Biforner + Wise-IoU and YOLOv7 + Biforner models have reduced parameter counts compared to YOLOv7, with the YOLOv7 + Biforner + Wise-IoU model having the second-highest inference speed among the four models, at 22.42 FPS. Compared to the YOLOv5n and YOLOv5x models, the improved YOLOv7 + Biforner + Wise-IoU model shows a 20% and 12% higher precision, an 11% and 25% higher recall than YOLOv5, and a 7% and 2% higher mAP@.5.

**Table 2.** Comparison of detection performance for different methods.

Model	P (%)	R (%)	mAP@.5	FPS	Parameters/M
YOLOv5n	70	76	0.77	38.23	1.8
YOLOv5x	75	68	0.81	18.14	86.2
YOLOv7	78	82	0.80	21.83	36.5
YOLOv7-e6e	80	76	0.79	12.20	164.8
YOLOv7 + Biformer	76	83	0.81	24.21	33.5
YOLOv7 + biformer + wise-iou	84	85	0.83	22.42	33.5

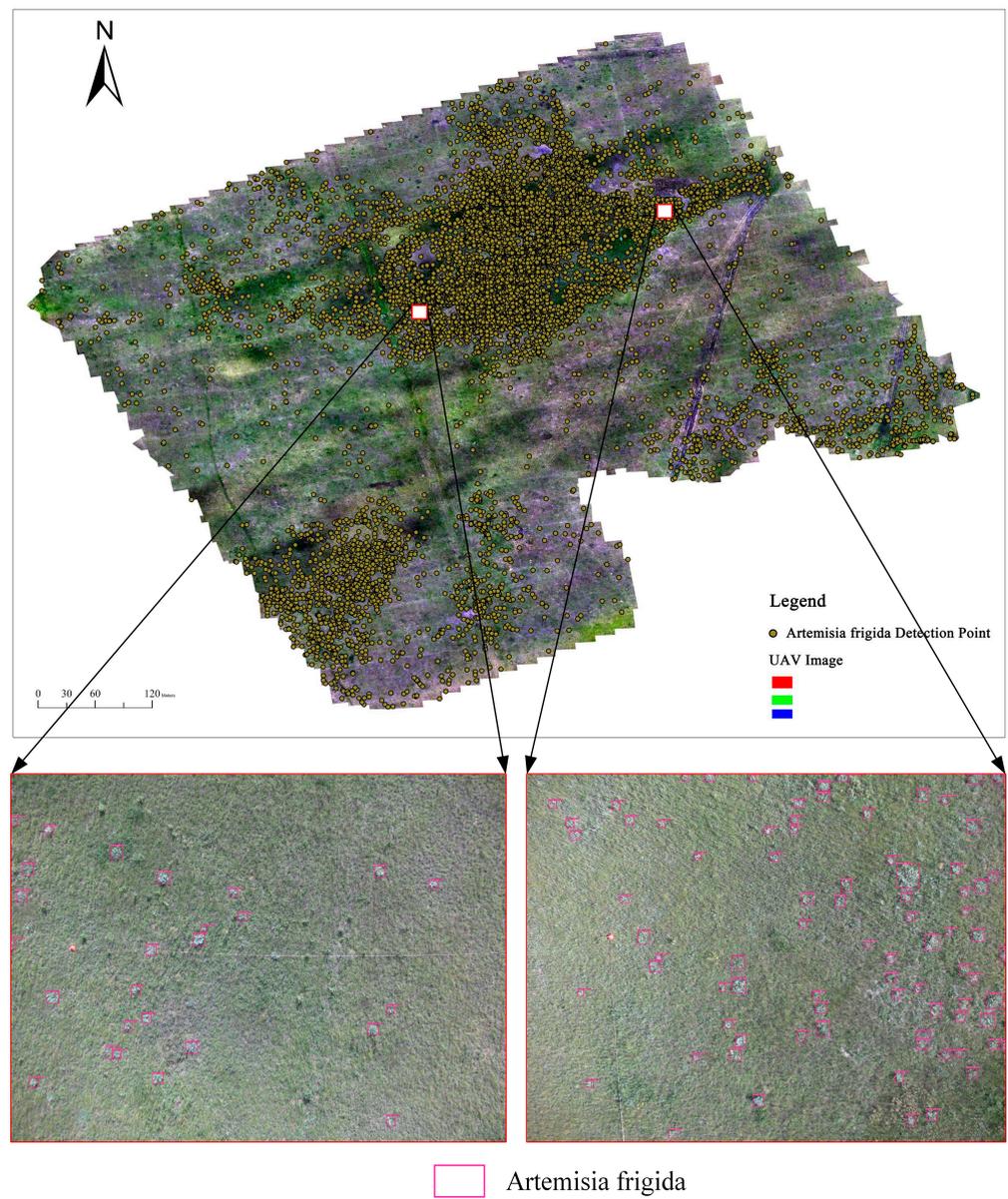
### 3.2. *Artemisia frigida* Detection and Spatial Distribution Mapping

We applied the improved YOLOv7 + biformer + wise-iou model to *Artemisia frigida* detection in a  $30 \times 30$  m plot using UAV images and compared the results with field survey results, as shown in Table 3. The overall precision for the 10 plots was 96%, and the recall was 94%, indicating superior overall performance of the model in *Artemisia frigida* detection at the plot scale. We applied the improved YOLOv7 (YOLOv7 + biformer + wise-iou) to the entire experimental area, obtaining the distribution points of *Artemisia frigida* across the entire area. Through ground truth selection and visual interpretation analysis, the detection accuracy was high, meeting the requirements for mapping the distribution of *Artemisia frigida*. After statistical analysis, a total of 15,663 *Artemisia frigida* distribution points were detected in the entire experimental area, as shown in Figure 9.

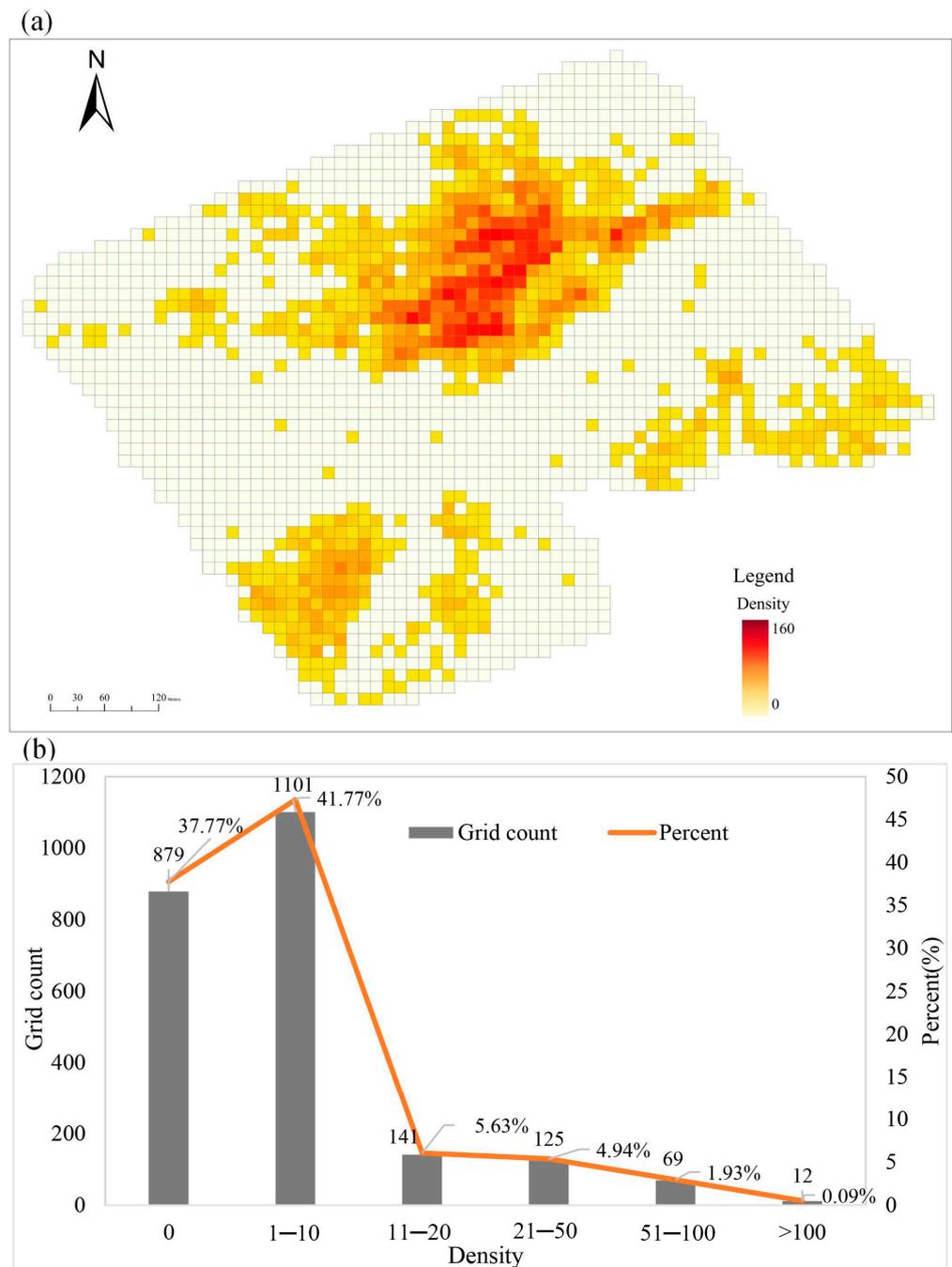
**Table 3.** The evaluation result of *Artemisia frigida* detection in ten plots.

Plot	TP	FP	FN	P (%)	R (%)
XM_01	0	0	0	0	0
XM_02	1	0	1	100	50
XM_03	6	0	2	100	75
XM_04	192	11	7	95	96
XM_05	68	4	7	94	91
XM_06	60	2	5	97	92
XM_07	116	3	6	97	95
XM_08	7	1	1	88	88
XM_09	36	0	1	100	97
XM_10	44	1	4	98	92
Overall	530	22	34	96	94

The distribution density of *Artemisia frigida* was obtained by counting detection points within each grid, as shown in Figure 10. In terms of density level, the average density of *Artemisia frigida* in the experimental area is 7 per 100 m<sup>2</sup>, with a maximum density of 160 per 100 m<sup>2</sup>. Among these, 879 grids have zero *Artemisia frigida* points, accounting for 33.77%. The grids with a density of 1–10 *Artemisia frigida* per 100 m<sup>2</sup> are 1101, representing 47.31% of the total. The grids with a density of 11–20 *Artemisia frigida* per 100 m<sup>2</sup> are 141, accounting for 6.06%. The grids with a density of 21–50 *Artemisia frigida* per 100 m<sup>2</sup> are 125, representing 5.37%. The grids with a density of 51–100 *Artemisia frigida* per 100 m<sup>2</sup> are 69, accounting for 2.97%. Finally, there are 12 grids with a density exceeding 100 *Artemisia frigida* per 100 m<sup>2</sup>, representing 0.52%.



**Figure 9.** The points of *Artemisia frigida* were obtained through object detection.



**Figure 10.** *Artemisia frigida* density. (a) *Artemisia frigida* density mapping; (b) Number of grids at different density levels.

#### 4. Discussion

Plant diversity can be measured at different scales, from satellite to ground level [24]. However, current remote sensing-based methods for observing plant diversity are mostly focused on the community or landscape scale and have not yet achieved individual-scale monitoring that matches ground-based observations [25]. With the development of near-ground remote sensing technologies such as UAV, high-resolution and ultra-high-resolution data can be obtained [26]. Combined with methods like deep learning, this allows for the rapid localization and classification of plants, greatly enhancing the efficiency of biodiversity surveys and monitoring [27–29].

In this study, the improvement of YOLOv7 model was implemented to enhance the detection performance of *Artemisia frigida* in UAV images, but there is still room for improvement in performance. Factors influencing the detection of grassland plants, including *Artemisia frigida*, mainly include the following. Firstly, it is necessary to obtain a sufficient amount of high-quality data. UAV can easily and quickly obtain images of the target area, but the resolution of the sensors and the environmental conditions during image acquisition may lead to some images not meeting the requirements for object detection data, ultimately affecting the detection performance [30,31]. Secondly, the quality of data annotation also affects the performance of the detection model. *Artemisia frigida* exhibits significant differences in morphology and size during different growth stages, making it challenging to distinguish the boundaries from the background, posing challenges for annotation [32]. Lastly, some *Artemisia frigida* has small bodies, and there may be overlapping and unclear boundaries between individuals, which will affect the final detection results.

Therefore, for the detection of dense small targets such as grassland plants, improving the algorithm can to some extent enhance the model's detection performance. However, data quality and sample quality directly impact the improved model's detection performance. Therefore, collecting high-quality datasets and building high-quality samples are essential for improving model accuracy. Currently, with the use of high-definition cameras equipped on UAV, analyzing the actual characteristics of the detection targets and planning flight routes and parameter settings reasonably can yield a sufficient number of high-quality samples. When combined with high-quality model improvements, this approach will greatly enhance the detection performance of grassland plants [33].

## 5. Conclusions

In this study, based on UAV imagery data, we conducted the detection of the grassland degradation indicator species, *Artemisia frigida*, by improving the YOLOv7 model. The results indicate that the improved YOLOv7 + Biformer + Wise-IoU model outperforms YOLOv7 overall, with a 6% improvement in precision and a 3% increase in mAP@0.5 compared to YOLOv7. The model has a reduced parameter count compared to YOLOv7, and its inference speed is close to that of YOLOv7. The YOLOv7 + Biformer + Wise-IoU model achieves an overall detection precision of 96% and a recall of 94% across 10 sample sites, demonstrating superior performance in the detection of *Artemisia frigida* at the site scale. Applying the model to predict *Artemisia frigida* across the entire experimental area and obtaining the density, the average density of *Artemisia frigida* in the entire experimental area is 7 per 100 m<sup>2</sup>, with a maximum density of 160 per 100 m<sup>2</sup>.

From the perspective of the model's detection performance on *Artemisia frigida* and its application in the experimental area, fully leveraging the advantages of UAVs in terms of agility, flexibility, and high-resolution data acquisition, combined with the application scenarios of grassland plant recognition, developing or improving adapted high-quality object detection models holds broad prospects in grassland species distribution surveys. However, there is still significant room for improvement in the model's detection accuracy and speed. Continuous iterations and optimizations are necessary to enhance the overall performance of the model, thereby expanding its applicable scenarios and deployable environments. Looking ahead to the requirements of future grassland surveys regarding the number of identifiable species and recognition accuracy, the research will focus on collecting UAV image data containing more grassland degradation indicator species and constructing high-quality UAV image samples of degradation indicator species. This will improve the model's detection performance for degradation indicator species and even more grassland species, establishing a comprehensive solution encompassing UAV flight control, sensor selection, data acquisition, data processing, model training, model deployment, and application.

**Author Contributions:** Conceptualization, Y.W., H.W. and J.G.; methodology, Y.W. and Z.H.; software, C.S. and B.Y.; validation, Y.W. and Z.H.; formal analysis, Y.W.; investigation, C.S.; resources, B.Y. and C.S.; data curation, Y.W. and B.Y.; writing—original draft preparation, Y.W.; writing—review and

editing, H.W. and J.G.; visualization, Y.W.; supervision, H.W. and J.G.; project administration, H.W.; funding acquisition, J.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by National Key R&D Program of China, grant number 2021YFB3901102.

**Data Availability Statement:** The raw data supporting the conclusions of this article will be made available by the authors on request.

**Conflicts of Interest:** Authors Huawei Wan, Jixi Gao, Chenxi Sun and Bin Yang was employed by the institution. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

- Liu, H.; Lu, W.; Chen, C. Research Progress of Grassland Degraded Succession and Diagnosis. *Acta Ecol. Sinica* **2011**, *19*, 865–871.
- Zhang, Y.; Zhu, J.; Shen, R.; Wang, L. Research progress on the effects of grazing on grassland ecosystem. *Chin. J. Plant Ecol.* **2020**, *44*, 553–564. [[CrossRef](#)]
- Wang, Z.; Jiang, L.; Wang, S.; Wang, Y.; Zhou, H. Assessment methods for grassland restoration. *Acta Ecol. Sinica* **2022**, *42*, 6464–6473.
- Wang, Z.; Wang, Z.; Lü, S.; Yan, B.; Wang, Z.; Men, X.; Bao, Y. Effects of stocking rate on population density and spatial distribution of *Artemisia frigida* in desert steppe. *Acta Ecol. Sinica* **2022**, *42*, 3420–3428.
- Liu, Z.; Li, Z. Plant biodiversity of *Aretemisia frigida* communities on degraded grasslands under different grazing intensities after thirteen-year enclosure. *Acta Ecol. Sinica* **2006**, *26*, 475–482.
- Jiang, Y.; Li, C.; Xu, R.; Sun, S.; Robertson, J.S.; Paterson, A.H. DeepFlower: A deep learning-based approach to characterize flowering patterns of cotton plants in the field. *Plant Methods* **2020**, *16*, 156. [[CrossRef](#)] [[PubMed](#)]
- Gai, Y.; Fan, W.; Xu, X.; Zhang, Y. Flower species identification and coverage estimation based on hyperspectral remote sensing data. In Proceedings of the 2011 IEEE International Geoscience and Remote Sensing Symposium, Vancouver, BC, Canada, 24–29 July 2011; pp. 1243–1246.
- Niu, Y.; Liu, T.; Duan, L.; Luo, Y.; Qi, X.; Chen, X. Analysis of Hyperspectral Characteristics and Extraction of Remote Sensing Interpreting Parameters of Five Types of Typical Psammo-Vegetation of the Horqin Sandy Land. *J. Ecol. Rural Environ.* **2017**, *33*, 632–644.
- Guo, Q.; Hu, T.; Liu, J.; Jin, S.; Xiao, Q.; Yang, G.; Gao, X.; Xu, Q.; Xie, P.; Peng, C.; et al. Advances in light weight unmanned aerial vehicle remote sensing and major industrial applications. *Prog. Geogr.* **2021**, *40*, 1550–1569. [[CrossRef](#)]
- Li, F. Application and Discussion of UAV Technology in Ecological Remote Sensing Monitoring of Grassland. *Bull. Surv. Mapp.* **2017**, *7*, 99.
- Gao, J.; Sun, F.; Huo, F.; Zhang, L.; Zhou, S.; Yang, T.; Bianba, Z. Application and Evaluation of Unmanned Aerial Vehicle Remote Sensing in Grassland Animal and Plant Monitoring. *Acta Agrestia Sinica* **2021**, *29*, 1–9.
- Yang, H.; Du, J. Classification of desert steppe species based on unmanned aerial vehicle hyperspectral remote sensing and continuum removal vegetation indices. *Optik* **2021**, *247*, 167877. [[CrossRef](#)]
- Gallmann, J.; Schüpbach, B.; Jacot, K.; Albrecht, M.; Winizki, J.; Kirchgessner, N.; Aasen, H. Flower mapping in grasslands with drones and deep learning. *Front. Plant Sci.* **2022**, *12*, 3304. [[CrossRef](#)] [[PubMed](#)]
- Eikelboom, J.A.; Wind, J.; van de Ven, E.; Kenana, L.M.; Schroder, B.; de Knegt, H.J.; Prins, H.H. Improving the precision and accuracy of animal population estimates with aerial image object detection. *Methods Ecol. Evol.* **2019**, *10*, 1875–1887. [[CrossRef](#)]
- Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
- Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7263–7271.
- Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
- Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
- Jiang, K.; Xie, T.; Yan, R.; Wen, X.; Li, D.; Jiang, H.; Wang, J. An Attention Mechanism-Improved YOLOv7 Object Detection Algorithm for Hemp Duck Count Estimation. *Agriculture* **2022**, *12*, 1659. [[CrossRef](#)]
- Zhu, L.; Wang, X.; Ke, Z.; Zhang, W.; Lau, R. BiFormer: Vision Transformer with Bi-Level Routing Attentio. *arXiv* **2023**, arXiv:2303.08810.
- Yang, Z.; Feng, H.; Ruan, Y.; Weng, X. Tea Tree Pest Detection Algorithm Based on Improved Yolov7-Tiny. *Agriculture* **2023**, *13*, 1031. [[CrossRef](#)]
- Zhang, Y.F.; Ren, W.; Zhang, Z.; Jia, Z.; Wang, L.; Tan, T. Focal and efficient IOU loss for accurate bounding box regression. *Neurocomputing* **2022**, *506*, 146–157. [[CrossRef](#)]
- Tong, Z.; Chen, Y.; Xu, Z.; Yu, R. Wise-IOU: Bounding Box Regression Loss with Dynamic Focusing Mechanism. *arXiv* **2023**, arXiv:2301.10051.
- Wang, R.; Gamon, J.A. Remote sensing of terrestrial plant biodiversity. *Remote Sens. Environ.* **2019**, *231*, 111218. [[CrossRef](#)]

25. Gamon, J.A.; Wang, R.; Gholizadeh, H.; Zutta, B.; Townsend, P.A.; Cavender-Bares, J. Consideration of scale in remote sensing of biodiversity. In *Remote Sensing of Plant Biodiversity*; Springer: Gewerbestrasse, Switzerland, 2020; pp. 425–447.
26. Lyu, X.; Li, X.; Dang, D.; Dou, H.; Wang, K.; Lou, A. Unmanned aerial vehicle (UAV) remote sensing in grassland ecosystem monitoring: A systematic review. *Remote Sens.* **2022**, *14*, 1096. [[CrossRef](#)]
27. Pi, W.; Du, J.; Bi, Y.; Gao, X.; Zhu, X. 3D-CNN based UAV hyperspectral imagery for grassland degradation indicator ground object classification research. *Ecol. Inform.* **2021**, *62*, 101278. [[CrossRef](#)]
28. Zhang, Y.; Du, J.; Pi, W.; Gao, X.; Wang, Y. Deep learning classification of grassland desertification in China via low-altitude UAV hyperspectral remote sensing. *Spectroscopy* **2022**, *37*, 28–35. [[CrossRef](#)]
29. Pöttker, M.; Kiehl, K.; Jarmer, T.; Trautz, D. Convolutional Neural Network Maps Plant Communities in Semi-Natural Grasslands Using Multispectral Unmanned Aerial Vehicle Imagery. *Remote Sens.* **2023**, *15*, 1945. [[CrossRef](#)]
30. Ramachandran, A.; Sangaiah, A.K. A review on object detection in unmanned aerial vehicle surveillance. *Int. J. Cogn. Comput. Eng.* **2021**, *2*, 215–228. [[CrossRef](#)]
31. Zhang, H.; Shao, F.; He, X.; Zhang, Z.; Cai, Y.; Bi, S. Research on Object Detection and Recognition Method for UAV Aerial Images Based on Improved YOLOv5. *Drones* **2023**, *7*, 402. [[CrossRef](#)]
32. Kong, J.; Zhang, Z.; Zhang, J. Classification and identification of plant species based on multi-source remote sensing data: Research progress and prospect. *Biodivers. Sci.* **2019**, *27*, 796–812.
33. Bakacsy, L.; Tobak, Z.; van Leeuwen, B.; Szilassi, P.; Biró, C.; Szatmári, J. Drone-Based Identification and Monitoring of Two Invasive Alien Plant Species in Open Sand Grasslands by Six RGB Vegetation Indices. *Drones* **2023**, *7*, 207. [[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.