

RESEARCH

Open Access



Lightweight highland barley detection based on improved YOLOv5

Minghui Cai^{1,2}, Hui Deng³, Jianwei Cai^{1,2}, Weipeng Guo¹, Zhipeng Hu¹, Dongzheng Yu¹ and Houxi Zhang^{1*}

Abstract

Accurate and efficient assessment of highland barley (*Hordeum vulgare* L.) density is crucial for optimizing cultivation and management practices. However, challenges such as overlapping spikes in unmanned aerial vehicle (UAV) images and the computational requirements for high-resolution image analysis hinder real-time detection capabilities. To address these issues, this study proposes an improved lightweight YOLOv5 model for highland barley spike detection. We chose depthwise separable convolution (DSConv) and ghost convolution (GhostConv) for the backbone and neck networks, respectively, to reduce the parameter and computational complexity. In addition, the integration of convolutional block attention module (CBAM) enhances the model's ability to focus on target object in complex backgrounds. The results show that the improved YOLOv5 model has a significant improvement in detection performance. Precision and recall increased by 3.1% to 92.2% and 86.2%, respectively, with an F1 score of 0.892. The AP_{0.5} reaches 92.7% and 93.5% for highland barley in the growth and maturation stages, respectively, and the overall mAP_{0.5} improved to 93.1%. Compared to the baseline YOLOv5n model, the number of parameters and floating-point operations (FLOPs) were reduced by 70.6% and 75.6%, respectively, enabling lightweight deployment without compromising accuracy. In addition, the proposed model outperformed mainstream object detection algorithms such as Faster R-CNN, Mask R-CNN, RetinaNet, YOLOv7, and YOLOv8, in terms of detection accuracy and computational efficiency. Although this study also suffers from limitations such as insufficient generalization under varying lighting conditions and reliance on rectangular annotations, it provides valuable support and reference for the development of real-time highland barley spike detection systems, which can help to improve agricultural management.

Keywords Object detection, YOLOv5, Lightweight, Highland barley, UAV

Introduction

Highland barley (*Hordeum vulgare* L.), as a crucial cereal crop, is widely cultivated in high-altitude regions worldwide due to its resistance to cold temperatures, broad

adaptability, high yield potential, and short growth cycles [1]. China is a major producer, with approximately 270,000 hectares dedicated to highland barley cultivation, mainly in the Xizang Autonomous Region [2]. In the Qinghai-Xizang Plateau, highland barley is the staple crop, accounting for 54.67% of the planted area and constituting 70.25% of the total grain production in 2022 [3]. Therefore, accurate monitoring of highland barley growth parameters, particularly plant density, is critical for effective crop management and yield prediction, promoting precision agriculture practices and ensuring food security in these challenging environments [4].

Traditional crop yield estimation methods primarily rely on destructive sampling or manual assessment,

*Correspondence:

Houxi Zhang

zhanghouxi@fafu.edu.cn

¹ College of Forestry, Fujian Agriculture and Forestry University, Fuzhou 350002, Fujian, China

² College of JunCao Science and Ecology, Fujian Agriculture and Forestry University, Fuzhou 350002, Fujian, China

³ College of Geography and Planning, Chengdu University of Technology, Chengdu 610059, Sichuan, China



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

which are time-consuming, labor-intensive, prone to subjectivity, and unrepresentative [5, 6]. As a result, they are difficult to apply to large-scale yield predictions. Therefore, there is an urgent need to explore an efficient, accurate, and labor-saving method for large-scale crop yield estimation. In recent years, remote sensing technology has provided a new method for estimating large-scale crop yield estimation [7]. However, traditional remote sensing technologies (e.g., satellite-based and aerospace-based) suffer from low spatial resolution, making it difficult to identify small objects like highland barley. Moreover, they are susceptible to weather factors like clouds and fog [8]. In contrast, unmanned aerial vehicle (UAV), with its merit of high spatial and temporal resolution, flexibility, and low cost [9], provides a new technical means for efficient and accurate detection of highland barley. However, higher-resolution image data typically represents more pixels and details, which requires larger storage space and more sophisticated processing power, placing higher demands on algorithms.

Object detection techniques based on computer vision have achieved promising results in agriculture with the progress of deep learning [10]. Deep learning based object detection typically offers better adaptability, faster speed, and higher accuracy than traditional algorithms [11]. Combining UAV technology, the deep learning based object detection provides an effective solution for accurate large-scale crop yield estimation. Currently, one-stage and two-stage algorithms are the most primary forms of deep learning based object detection. Two-stage object detection algorithms, such as Faster R-CNN and Mask R-CNN, have established themselves as effective approaches. However, in the realm of real-time performance, one-stage object detection algorithms, represented by the YOLO series, have significantly surpassed their two-stage counterparts [12]. The YOLO algorithm dominates the market due to its unparalleled efficiency, commendable accuracy and simplified training procedures, making it the first choice for many applications [13].

Traditional object detection algorithms often sacrifice detection speed for higher detection accuracy, which results in computationally demanding models that struggle to meet real-time requirements [14]. However, detection speed is one of the most important performance indicators for object detection [15]. Meanwhile, real-time detection is fundamental for practical applications. In the field of object detection, since the YOLO algorithm is highly efficient, modularity, and easy to improve, it has received considerable attention.

From YOLOv1 to YOLOv10, its basic and improved models have been widely used in various domains.

Mendes et al. [16] proposed that YOLOv5 stands out as the superior real-time object detection model due to its exceptional inference speed, precision, low training time. Yang et al. [17] also proposed that the YOLOv5 algorithm has the advantages of high accuracy, speed and performance. In summary, YOLOv5 has become a popular choice in the field of real-time detection due to its higher detection accuracy, smaller parameter number and floating point operations (FLOPs), and ease of lightweight implementation. For example, Yao et al. [18] realized real-time detection of kiwifruit defects based on an improved lightweight YOLOv5 model. Chen et al. [19] developed a real-time strawberry diseases detection algorithm, providing a new way to identify and control strawberry disease. Yu et al. [20] devised a real-time detection of pineapple flowers based on an improved lightweight YOLOv5 model. These research results demonstrate the significant advantages of YOLOv5 in terms of detection effect and lightweight implementation, as well as its vast potential for real-time application.

As the vital food crop in the Xizang Autonomous Region of China [21], accurate estimation of the yield of highland barley is significant to guarantee food security in China and globally. However, it is difficult to estimate highland barley yield by traditional methods, which face challenges in accurately estimating highland barley yield, and existing research on crop yield estimation based on object detection still pays limited attention to highland barley. Therefore, it is essential to explore the role of object detection technology in highland barley yield estimation and management. The highland barley spike, a key component of the plant, exhibits distinct color and morphological characteristics compared to other parts, making it a reliable indicator to identify highland barley. Moreover, because the highland barley harvest time varies in different growth periods, assessing the growth period of highland barley can provide more detailed yield estimation and management of highland barley.

Therefore, we propose a lightweight model based on the improved YOLOv5 algorithm for the detection of highland barley spike. Aiming to meet the demands for real-time detection, simplicity and lightweight are achieved while ensuring detection accuracy. The effectiveness of the proposed method is verified by comparative analysis with existing object detection algorithms. This study provides an efficient new method for real-time detection, and precise management of highland barley, and promotes the intelligent and modern development of the highland barley industry.

Materials and methods

Study area

The study area is located in Sangzhuzi District ($82^{\circ}00'E - 90^{\circ}20'E$, $27^{\circ}23'N - 31^{\circ}49'N$) and Lazi County ($87^{\circ}24'E - 88^{\circ}21'E$, $28^{\circ}47'N - 29^{\circ}37'N$) in Rikaze City, Xizang Autonomous Region, China (Fig. 1). Rikaze belongs to the typical Qinghai-Xizang Plateau region, characterized by a plateau subfrigid semi-arid climate, with an average annual precipitation of 400 mm, an annual temperature of $6.3^{\circ}C$, and an average altitude above 4,000 m. The region has thin air, and low air pressure, low air oxygen content, but sunny, strong ultraviolet rays, long sunshine hours, and an average annual sunshine time of 3300 h, providing a unique condition for highland barley growth. As one of the most widely planted highland barley regions in China, Rikaze City is known as the “home

of highland barley in the world”. According to the Department of Agriculture and Rural Affairs of Xizang Autonomous Region statistics, Rikaze City has a highland barley planting area of more than 60000 hectares, with an output of 408,900 tons, accounting for 49.13% of the region's total output in 2022 [22]. This underscores its significance as a vital “granary of Xizang”.

Figure 1b shows the Quxia Township of Lazi County. Quxia Township is mainly agricultural and situated within the Yarlung Zangbo River Basin, which is rich in water, flat, and has good soil texture, providing ideal natural conditions for crop growth. Figure 1d shows Nieixiong Township in Sangzhuzi District. Sangzhuzi District has sufficient water sources, which are suitable for the growth of highland barley and other alpine crops, and is the main agricultural area of Rikaze. Among them,

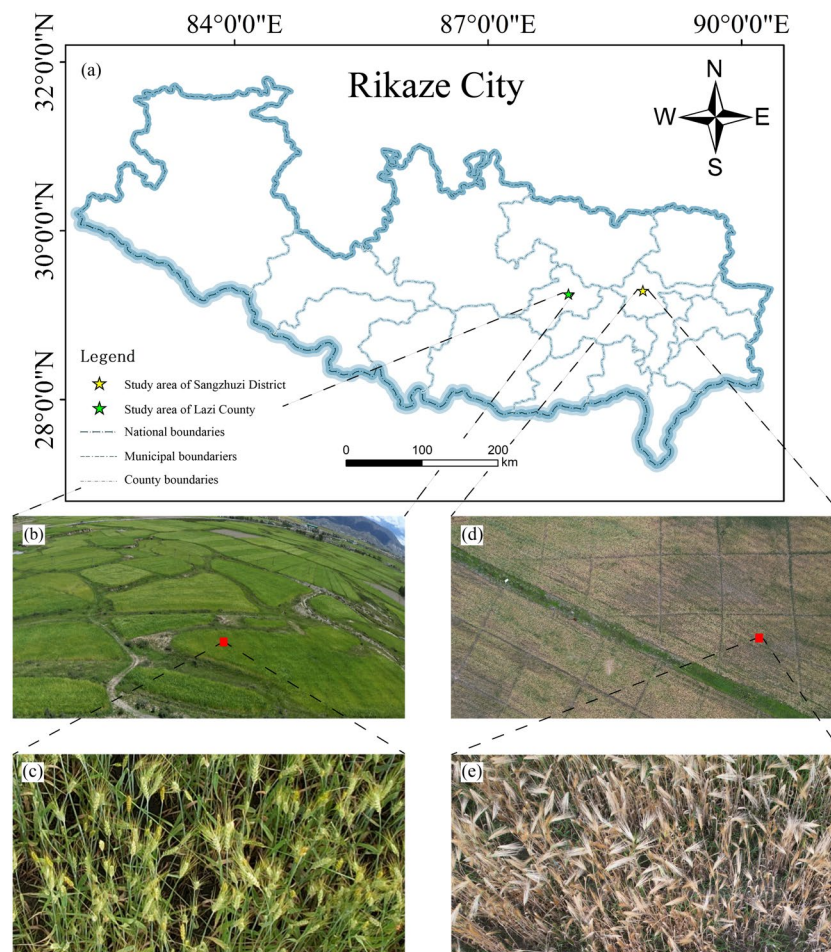


Fig. 1 Study area in Rikaze City, Xizang Autonomous Region. **a** Geographic location of the study area. **b** Overview of the highland barley field during the growth stage. **c** Details of highland barley plants during the growth stage. **d** Overview of the highland barley during the maturation stage. **e** Details of highland barley during the maturation stage. The Rikaze City standard map is downloaded from <http://zrzyt.xizang.gov.cn/fw/zyxz/202004/P020240716645043257282.jpg> (ZangS (2024) 034, approved by the Department of Natural Resources of Xizang Autonomous Region), and the base map has not been modified

Nierixiong Township is known as “the land of plenty”, with abundant water resources and fertile soil, which is suitable for agricultural development. These two locations have extensive highland barley cultivation areas and abundant yields, making them ideal sites for conducting highland barley object detection research.

Overall workflow

In this study, we propose a deep learning based object detection algorithm to detect and count highland barley spikes in UAV-based images, and the overall workflow is shown in Fig. 2. (1) Data acquisition: image acquisition using a UAV with a visible light sensor. (2) Data preprocessing: using a simple code for automatic cropping to crop the original image into uniform-sized subgraphs, avoiding the fuzzy edge areas, and screening the middle part of the subgraphs with higher quality for the study. (3) Dataset building: using LabelImg software (version 1.8.6) to label the highland barley in the screened image, ensuring that each highland barley spike is located in the center of the bounding box, and producing a highland barley image dataset. (4) Model evaluation: comparing the accuracy of different object detection algorithms for the application of highland barley spike recognition and the size and complexity of the algorithms to validate this study's proposed model's complexity to substantiate the performance merits of our proposed methodology in this study.

Dataset generation

Data acquisition and preprocessing

To comprehensively assess the impact of highland barley growth stages on model recognition accuracy, this study collected highland barley image data in different periods and locations. Specifically, the growth highland barley

images were acquired on August 8, 2022, in Lazi County, and the transition and maturation highland barley images were acquired on August 23, 2023, in Sangzhu District. These two dates were chosen based on the fact that highland barley in the Rikaze region is usually harvested in September, when a small amount of growing highland barley and a large amount of mature highland barley exist in the field at the same time, providing effective samples for the model to distinguish between different growth stages.

In this study, we used a DJI Air 2S UAV to collect data, which was equipped with a 1-inch 20-megapixel visible sensor (pixel size: $2.4\mu\text{m}$, equivalent focal length: 22 mm), enabling to effectively capture the fine features of highland barley in the field. Finally, we acquired 501 RGB images with a resolution of 5472×3648 pixels. Since the dense distribution of highland barley, the small size spikes, and the edges of the UAV aerial images frequently appear blurred. Therefore, it is difficult to directly use the original images for labeling. To enhance the efficiency and accuracy of data annotation, we adopt the following preprocessing process. (1) image segmentation: each aerial image was divided into 35 equal parts ($5 \text{ rows} \times 7 \text{ columns}$) using code; (2) regional screening: selecting the segmented 3×3 subimages located in the middle region of a total of 9 subimages to be annotated, which effectively avoids the edge of the fuzzy region and ensures the quality of the annotated data (Fig. 2).

Data annotation, augmentation and partitioning

In this study, we use LabelImg software to annotate the selected highland barley images, ensuring that each highland barley spike was centered within its bounding box. To achieve more accurate yield estimation and management, the growth stages of highland barley were

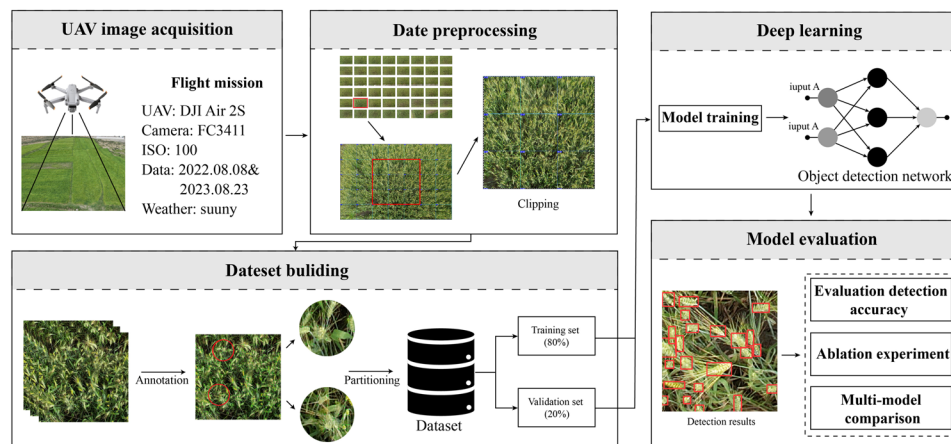


Fig. 2 Overall workflow

distinguished based on the color differences of the spikes: highland barley spikes that were green were labeled as growth period, while those that were golden yellow were labeled as maturation period. This approach aims to enable the model to differentiate between the various growth stages of highland barley, and thereby improve the accuracy of yield estimation and management.

Deep learning, as a data-driven technique, requires a large number of training samples in order to learn the multi-level features of the target for better detection [23]. Data augmentation can effectively improve the robustness and adaptability of the model by increasing the diversity of samples [24]. In this study, data enhancement is performed on the original dataset using geometric distortion, adding noise, occlusion, and sharpening, as shown in Fig. 3. Among them, geometric distortion uses random rotation ($0^\circ, 90^\circ, 180^\circ, 270^\circ$) and flip to strengthen the model's detection ability to diverse angles. Meanwhile, we also added noise, occlusion, and sharpening to boost the model's performance to recognize highland barley in complex environments.

Before data augmentation, 135 images were labeled, which were then processed to create the final dataset containing a total of 2970 images. To train and validate the performance of the object detection model, the labeled data are divided into a training set and a validation set in the ratio of 8:2 to construct the highland barley spikes recognition dataset.

Depthwise separable convolution

To improve the computational effectiveness of the model, this study introduces depthwise separable convolution (DSConv) into the neck network of the YOLOv5 model. The standard convolution operation is decomposed by DSConv into depthwise convolution and pointwise convolution. This decomposition reduces the number of parameters required for convolutional computations and effectively improves the utilization efficiency of kernel parameters [25]. As shown in Fig. 4, depthwise convolution is performed first, applying independent convolution operations to every individual channel to the input feature map.

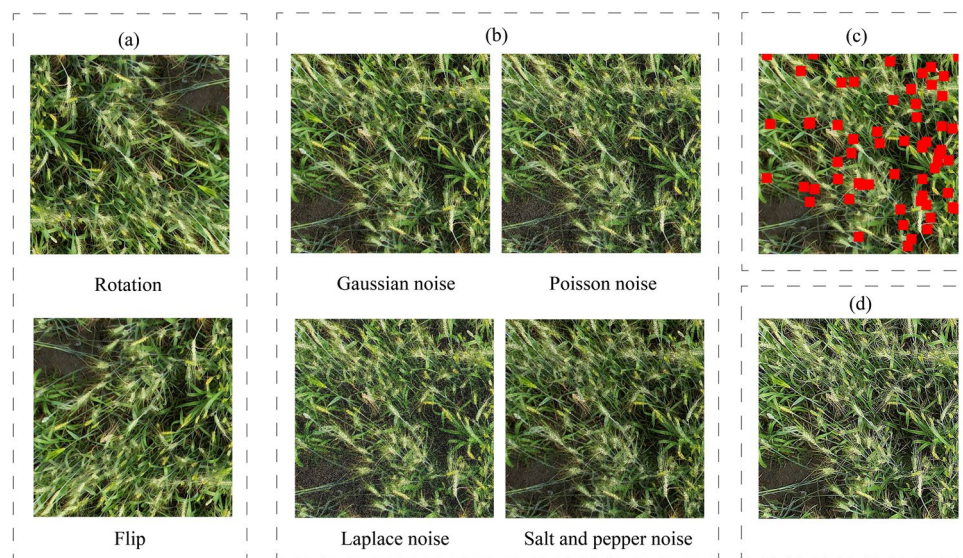


Fig. 3 Data augmentation. **a** Geometric distortion. **b** Noise. **c** Occlusion. **d** Sharpening

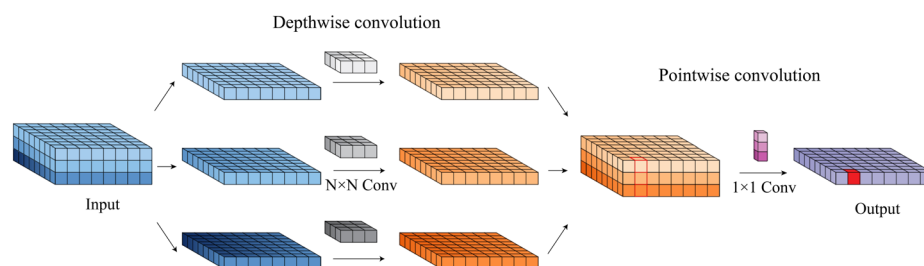


Fig. 4 Depthwise separable convolution

Subsequently, pointwise convolution fuses the feature maps of all channels, generating the final output feature map.

The parameters are calculated based on the kernel size ($k \times k$), the number of input channels (in channels), and the number of output channels (out channels). The number of parameters for traditional convolution is given by the Eq. 1.

$$\text{Parameters} = k \times k \times \text{in channels} \times \text{out channels} \quad (1)$$

For DSConv, the depthwise convolution, independent convolution operations are applied to every individual channel of the input feature map. The number of parameters for depthwise convolution is given by the Eq. 2. Subsequently, pointwise convolution uses a 1×1 convolution to fuse the feature maps of all channels, generating the final output feature map. The number of parameters for pointwise convolution is given by the Eq. 3.

$$\text{Parameters} = k \times k \times \text{in channels} \quad (2)$$

$$\text{Parameters} = 1 \times 1 \times \text{in channels} \times \text{out channels} \quad (3)$$

Compared to the traditional convolution, the DSConv significantly decreases the quantity of parameters required for convolution calculations by separating the correlations between spatial dimensions and channel dimensions. This reduction maintains the network's performance while lowering computational complexity, facilitating model lightweight, and making it easier to deploy on resource-constrained devices.

Ghost convolution

To further reduce model parameters and computational burden, this study introduces ghost convolution (GhostConv) into the YOLOv5 model to replace standard convolution. The traditional convolution operation maps all channels of the input feature map with convolution, which is prone to produce redundant features and requires a large amount of parameter and computational resources [26]. Ghost convolution extend intrinsic feature maps through simple linear operations, such as unit mapping and linear transformation. [27]. The network structures of traditional convolution and GhostConv are shown in Figs. 5 and 6, respectively.

Compared to traditional convolution, GhostConv generates additional feature maps through simple linear operations, significantly reducing the model's computational complexity while maintaining high detection accuracy. This characteristic makes it widely applicable in lightweight models.

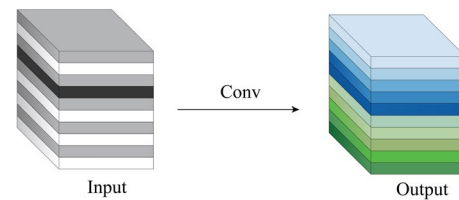


Fig. 5 General convolution

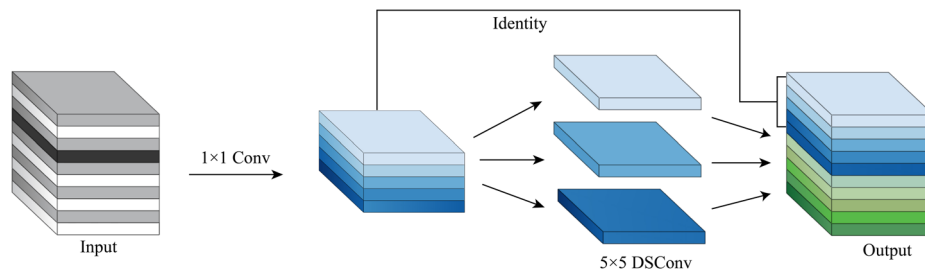
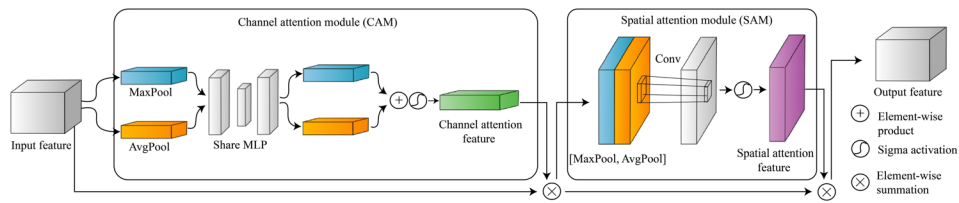
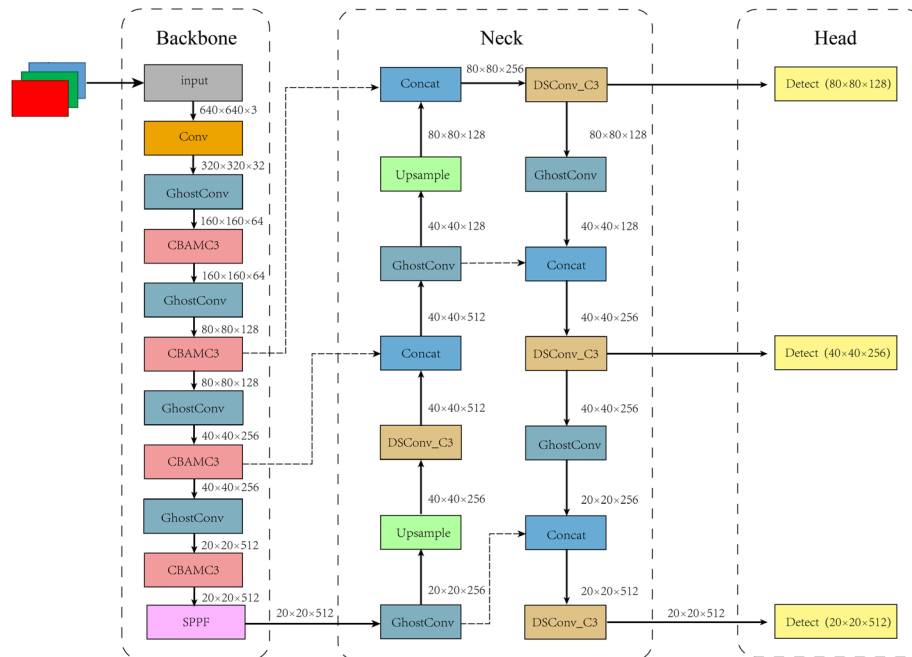
Convolutional block attention module

This study introduces a convolutional block attention module (CBAM) into the YOLOv5 model to replace the C3 module to further strengthen the detection performance of the object detection network. CBAM is a lightweight attention mechanism module that can direct attention to important feature information to improve the model's recognition ability [28]. CBAM incorporates the channel attention module (CAM) and the spatial attention module (SAM), as shown in Fig. 7 [29]. The CAM focuses on identifying the crucial weights of individual channels, enhancing the important feature channels and suppressing the unimportant feature channels. The SAM aims to discern the important weight of each spatial location and highlight important feature regions. This dual attention mechanism effectively improves the image recognition performance of the model.

Improved YOLOv5 network architecture

YOLOv5 is an efficient one-stage object detection model. Its detector primarily consists of a backbone network, a neck network, and a head network [30]. Based on network depth and width, YOLOv5 models can be categorized into YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. With the smallest storage space requirement, the fewest parameters, and the lowest FLOPs, the YOLOv5n model is suitable as a baseline for lightweight object detection tasks.

Therefore, this study adopted the YOLOv5n model as the baseline and implemented modifications to its network structure to enhance its performance in highland barley spike detection. The improved YOLOv5 network structure is shown in Fig. 8. Firstly, DSConv module is introduced into the neck network to reduce the number of parameters required for convolutional computation, and thus improve network computational efficiency. Secondly, GhostConv module is introduced into the backbone network to further decrease the model's parameter count and computational complexity. Lastly, CBAM module is incorporated into the neck and backbone networks to improve the ability of the model to pay attention to the key objects in complex backgrounds improving the recognition accuracy of the model.

**Fig. 6** Ghost convolution**Fig. 7** Convolutional block attention module**Fig. 8** Improved YOLO network structure diagram

Performance metrics

In this study, we select the following metrics as indicators to evaluate the performance of the object detection model. Precision (P) and recall (R) are important indicators for the evaluation of object detection performance, and the higher values of P and R indicate the better detection ability of the model, as shown in Eqs. 4 to 5.

$$P = \frac{TP}{TP + FP} \quad (4)$$

$$R = \frac{TP}{TP + FN} \quad (5)$$

where TP, FP, and FN respectively represent the number of correctly detected highland barley (true positives), the

number of falsely detected highland barley (false positives), and the number of missed detections of highland barley (false negatives).

The F1 score is calculated as the harmonic mean of P and R, providing a comprehensive evaluation of a model's performance, as shown in Eq. 6.

$$F1 = 2 \times \frac{P \times R}{P + R} \quad (6)$$

The average precision (AP) is the area under the P and R curve. This metric measures the model's sensitivity to objects and reflects the overall performance of the model, as shown in Eq. 7. The mean average precision (mAP) is the average of all APs, as shown in Eq. 8.

$$AP = \int_0^1 p_i(R) dR \quad (7)$$

$$mAP = \frac{1}{N} \sum_{i=1}^N \int_0^1 p_i(R) dR \quad (8)$$

where $p_i(R)$ denotes the PR curve plotted using precision and recall, and N is the number of categories.

In addition, FLOPs and the number of parameters are used as a measure of model complexity and lightweight effect.

Results

Comparison of lightweight models

In this study, in order to select lightweight modules suitable for highland barley spikes, the effectiveness of MobileNetv2 [31], MobileNetv3 [32] (available in both large and small versions), Shufflenetv2 [33], DSConv, and GhostConv in detecting highland barley spikes was compared, as shown in Table 1. Soviany et al. [34] pointed out that an important consideration in the object detection task is to balance the accuracy and speed of detection. Therefore, in order to balance these two factors, two modules, DSConv and GhostConv, are chosen for this experiment to achieve lightness. These two models not only

reduce the computation and FLOPs of the model, but also slightly improve the experimental accuracy, which is perfectly suited to the research purpose. Although other lightweight models perform well in lightweight, especially MobileNetv2 greatly reduces the size and computation of the model, it also inevitably reduces the accuracy of the model to a greater extent, making them unsuitable for the detection of highland barley spikes.

Model training results

The trend of accuracy metrics of the improved YOLOv5 model in the training process is shown in Fig. 9. In this study, the specific details of the experimental hardware and software environment are presented in Table 2. The results indicate that precision, recall, mAP, and F1 score improve gradually over 300 epochs until leveling off at 92.1%, 86.3%, 93.1%, and 0.89, respectively. Meanwhile, the loss value continues to decrease and approaches the minimum value of 0.23, which indicates that the model is well trained and converges to an optimal state.

Ablation experiments

In this study, the effect of different modules on the model performance was assessed by ablation experiments, as shown in Table 3. The results show that the final improved model M7 combining the DSConv, GhostConv and CBAM modules achieved the best performance: accuracy of 92.2%, recall of 86.2%, and an F1 score of 0.892. The AP_{0.5} for highland barley growth and maturation were 92.7% and 93.5%, respectively, and the mAP_{0.5} was 93.1%. And the baseline model YOLOv5n had an accuracy of 89.1%, a recall of 83.1%, and an F1 score of 0.860, mAP_{0.5} of 90.3%. Compared with the baseline model, M7 has an overall improvement in these metrics. In addition, the number of parameters and FLOPs have been reduced to 1.2 M and 4.1 G, respectively, which represent 70.6% and 75.6% of the baseline model. This indicates that the combined use of DSConv, GhostConv and CBAM modules can effectively improve the recognition accuracy of the model for highland barley spikes, and at

Table 1 Comparison of lightweight models

Model	P/%	R/%	F1	mAP _{0.5} /%	Parameter/10 ⁶	FLOPs /G
YOLOv5n	89.1	83.1	86.0	90.3	1.76	4.1
YOLOv5n+Mobilenetv2	66.3	60.0	63.0	63.6	0.21	0.8
YOLOv5n+Mobilenetv3_large	75.7	74.9	75.3	78.3	0.98	2.5
YOLOv5n+Mobilenetv3_small	68.9	70.5	69.7	72.5	0.45	0.9
YOLOv5n+Shufflenetv2	62.7	56.6	59.5	59.0	0.22	0.5
YOLOv5n+DSConv	90.7	85.0	87.8	92.1	1.51	3.6
YOLOv5n+GhostConv	92.7	86.3	89.4	93.2	1.46	3.6

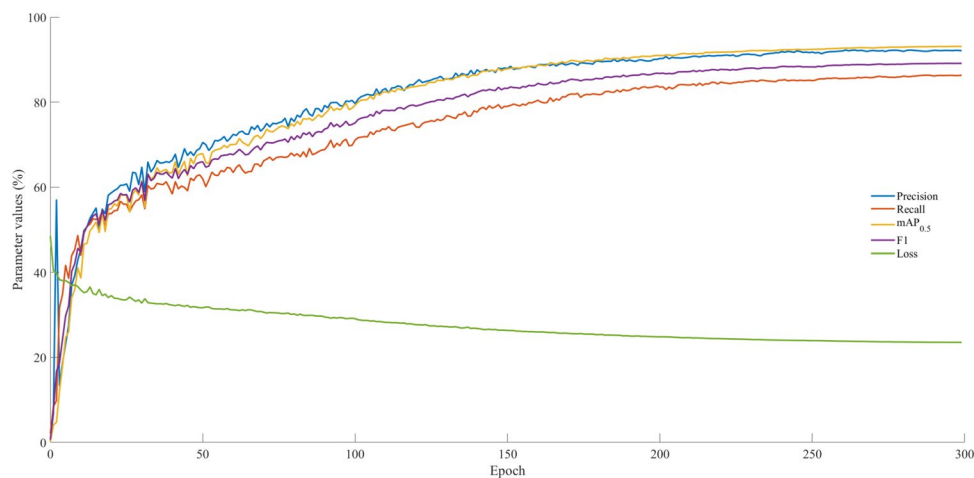


Fig. 9 Variation trend of different accuracy indicators during the model training

Table 2 Software and hardware environments

Item	Parameters and versions
Central processing unit (CPU)	AMD Ryzen 7 7840H w 8-core processor @ 3.8GHz
Memory (RAM)	64 GB
Solid state disk (SSD)	1TB
Graphics card (GPU)	NVIDIA RTX 4060 GPU (8GB)
Operating system (OS)	Windows 11
Programming environment (ENVS)	Torch 2.0.1 + CUDA11.7 + Python 3.11.4

Table 3 Result of ablation experiment

Model	DSConv	Ghost Conv	CBAM	P/%	R/%	F1	AP _{0.5} growth maturation	mAP _{0.5} /%	Parameters /10 ⁶	FLOPs /G	
YOLOv5n				89.1	83.1	0.860	90.7	89.9	90.3	1.7	4.1
M1	*			90.7	85.0	0.878	92.0	92.2	92.1	1.5	3.6
M2		*		92.7	86.5	0.895	93.1	93.3	93.2	1.4	3.6
M3			*	93.1	89.8	0.914	94.8	95.3	95.0	1.7	4.2
M4	*	*		92.0	85.2	0.885	92.3	92.7	92.5	1.2	3.0
M5	*		*	91.5	86.4	0.889	92.6	93.4	93.0	1.5	3.6
M6		*	*	92.6	87.7	0.901	93.8	94.0	93.9	1.4	3.6
M7(Ours)	*	*	*	92.2 (+3.1)	86.2 (+3.1)	0.892 (+0.032)	92.7 (+2.0)	93.5 (+3.6)	93.1 (+2.8)	1.2 (70.6%)	3.1 (75.6%)

* means that the improvement factor was used

the same time achieve the lightweight of the model, making it more suitable for practical application scenarios.

For other improvement methods, introducing the DSConv or GhostConv module alone can achieve model lightweighting while maintaining or even improving model accuracy. Among them, the introduction of the GhostConv module (M2) has the most significant effect on decreasing the number of parameters and

calculations. The F1 score reached 0.895, and the number of parameters and FLOPs were reduced to 1.46 M and 3.6 G respectively. The introduction of the CBAM module (M3) significantly enhanced the model accuracy without increasing the quantity of parameters and FLOPs, and the F1 score reached 0.914. In order to verify the effectiveness of various modules, we have introduced the complex number module at the same time. M4, which introduces

the DSConv and GhostConv modules at the same time, reduces the parameter amount to 1.21 M, and the F1 score increases to 0.885. M5, incorporating both DSConv and CBAM modules, reduces parameters to 1.52 M and improves the F1 score to 0.889. M6, incorporating both GhostConv and CBAM modules, reduces parameters to 1.48 M and improves the F1 score to 0.901.

Comparison with different algorithms

To further verify the effectiveness of the improved model M7, this study compared it with the current mainstream two-stage and one-stage object detection algorithms, as shown in Table 4.

Table 4 Comparison of different object detection algorithms

Model	mAP _{0.5}	Parameters/10 ⁶	FLOPs/G
M7(Ours)	93.1	1.2	3.1
YOLOv5n	90.3	1.7	4.1
Faster R-CNN	70.6	13.9	14.0
Mask R-CNN	73.1	14.1	14.0
RetinaNet	67.8	11.7	55.6
YOLOv7	94.8	36.4	103.2
YOLOv8n	93.8	3.0	8.1

The M7 outperforms Faster R-CNN, Mask R-CNN, and RetinaNet in detection performance while maintaining the advantage of being lightweight. Compared with YOLOv7 and YOLOv8n, although the accuracy is slightly lower, it has significant advantages in the number of parameters and FLOPs, rendering it more appropriate for real-time detection scenarios.

Visualization comparison

Figure 10 shows the detection results of visualization comparison between the original YOLOv5n and the improved model. In this figure, the red border represents the highland barley spikes during the growth period, the blue border represents the highland barley spikes during the maturation period, the orange arrows represent missed detections, and the purple arrows represent false detections. During the growth period, YOLOv5n detected only 28 highland barley spikes, while M7 detected 41. In the transition period, both YOLOv5n and M7 detected 6 growth highland barley spikes and 2 maturation highland barley spikes, but YOLOv5n showed one missed and false detection. During the maturation period, YOLOv5n detected only 2 highland barley spikes while M7 detected 3. It indicates that the improved model outperforms the original YOLOv5n model in detecting small target highland barley spike particles.

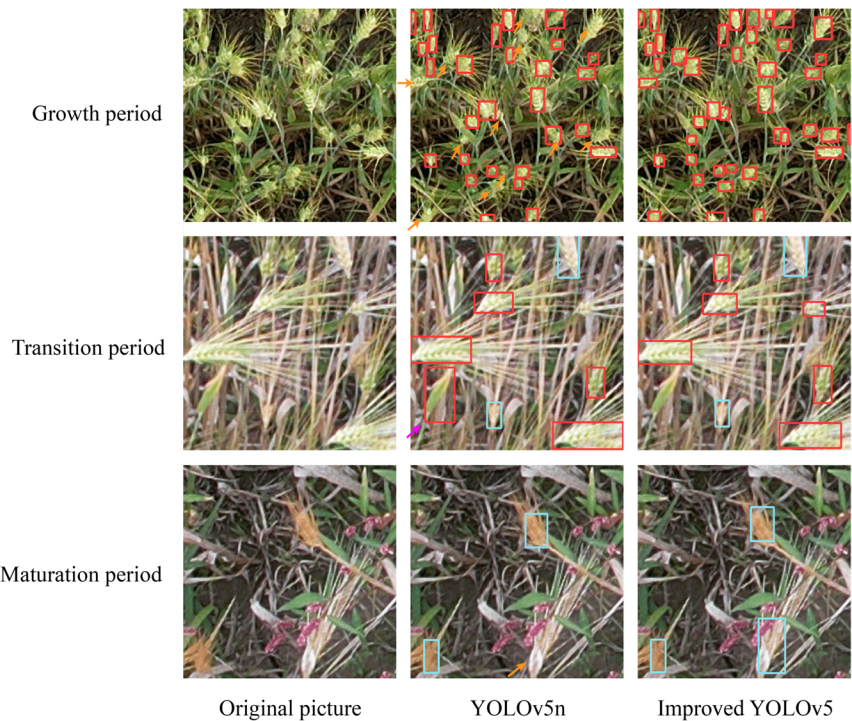


Fig. 10 Comparative analysis of highland barley spike detection using YOLOv5n and the improved YOLOv5 model. Red bounding boxes indicate highland barley spikes in the growth stage, while blue bounding boxes represent spikes in the maturation stage. Orange arrows highlight instances of missed detection by the YOLOv5n model, and purple arrows indicate false positive detections by the YOLOv5n model

Meanwhile, the improved model can maintain high recognition accuracy even in complex field environments, distinguish highland barley spikes with different growth states, and successfully complete the target detection task. This affords a new way for accurate real-time monitoring of highland barley spikes.

Discussion

Model performance comparison

This study addresses the challenge of accurately identifying highland barley spikes under field conditions, including partially shaded scenarios, by proposing a highland barley spike detection method based on an improved YOLOv5. By introducing DSConv, GhostConv, and CBAM modules, a lightweight and efficient highland barley spike detection model (M7) was successfully constructed. This model achieves accurate identification of highland barley spikes at different growth stages, which offers a new approach for real-time highland barley detection and rapid yield estimation. The results demonstrate that the M7 has achieved better performance in detection accuracy, model size, and computation efficiency. Compared with two-stage algorithms such as Faster R-CNN, Mask R-CNN and RetinaNet, the model has significant advantages in detection accuracy, the number of parameters and FLOPs. Compared with one-stage algorithms such as YOLOv7 and YOLOv8, the model also has significant advantages in the number of parameters and FLOPs, which is more suitable for real-time detection scenarios. This is mainly attributed to the efficient and lightweight design of the YOLOv5 network itself [35], as well as the lightweight modules and attention mechanism introduced in this study.

Recently, the series of YOLO algorithms have attracted much attention in the realm of object detection owing to their speed and efficiency. Among them, YOLOv5 has become one of the preferred solutions for lightweight object detection due to its flexible network structure and excellent performance. This study selected YOLOv5n as the base model and enhanced its performance by adding DSConv, GhostConv, and CBAM modules. Compared with the latest algorithms such as YOLOv7 [36] and YOLOv8 [37], the improved model has slightly lower mAP but shows significant advantages in lightweight design. YOLOv7 and YOLOv8 often adopt deeper and more complex network structures to achieve higher detection accuracy, which results in a substantial increase in model parameters and computational cost [38]. However, in practical applications, especially on resource-constrained mobile devices or embedded systems, model lightweight is particularly important. Therefore, the improved YOLOv5 model proposed in this

study achieves model lightweight while maintaining high detection accuracy, which makes it more practical.

Impact of model improvement strategies

To evaluate the influence of our lightweight strategies on model performance, this study employed the number of parameters and FLOPs as metrics, consistent with the metrics used by Zhang et al. [39] and Sun et al. [40]. To achieve model lightweight, this study uses DSConv and GhostConv to replace standard convolution operations. These two methods can effectively decrease the number of parameters and enhance detection efficiency while maintaining high detection accuracy, as validated by Yang et al. [41] and Chen et al. [42] in tomato and tea bud detection algorithms, respectively.

To address the difficulty of effectively extracting features from complex UAV aerial highland barley images using the original network structure, this study incorporates the CBAM attention mechanism. Zhang et al. [43] pointed out that attention mechanisms are a common method to enhance the feature extraction capability of YOLOv5. CBAM can capture the correlation between features in different dimensions, thereby improving the performance of image recognition tasks. In this study, the CBAM module was introduced to replace the standard C3 module, resulting in significant detection accuracy improvements. This has also been verified in the study of fresh tea bud identification by Guo et al. [44]. In addition, due to the overlap between highland barley spikes and the differences in individual growth, this study also adopts data augmentation to increase the number of training instances and suppress overfitting, thereby improving the model's generalization ability and robustness [45].

Limitations and future work

Although the improved YOLOv5 model has been enhanced in terms of parameters, complexity, and detection performance, some limitations remain. Firstly, since the data collection time is mostly concentrated at noon, the influence of different lighting environments on the object detection model is less considered. Although data augmentation can simulate images under different lighting conditions through optical transformation, this does not fundamentally solve the problem of insufficient generalization ability in diverse lighting environments [46]. Future work will focus on assessing more images in different lighting conditions to enrich the dataset and improve the model's transferability. Secondly, the rectangular annotation method used in this study may affect the detection accuracy. Future work could explore more refined

annotation methods, such as polygon or ellipse annotation [47]. Finally, Due to the presence of blurring around the original image pictures captured by the UAV and the small target of the barley spike, the blurring will have a relatively large impact on the detection results. Therefore, this part of the blurred data is not used in the dataset of this study, so it is difficult to detect the target in the original image captured by the UAV. So to address against this problem, this experiment designs a simple code for automatic cropping, which automatically crops, classifies and saves the images after acquiring the original images captured by the UAV. Then, the images to be detected in the specified folder are read for target detection. This method requires an extra step of processing the image, which is not conducive to real-time detection. In the future, better methods need to be investigated to achieve detection of small object for fuzzy images.

Conclusion

This study present a lightweight YOLOv5 model for the identification and counting of highland barley spikes in UAV aerial images. By introducing lightweight modules and an attention mechanism, the model achieves lightweight while maintaining high recognition accuracy. Experimental results demonstrate that the improved model is superior to other mainstream object detection algorithms in terms of recognition accuracy and counting accuracy. Notably, the final improved model (M7) achieves the best performance across all metrics: precision increased to 92.2%, recall increased to 86.2%, F1 score reached 0.892, and mAP_{0.5} increased to 93.1%. Compared to the baseline model, the number of parameters and computations are reduced by 70.6% and 75.6%, respectively. These improvements make the model particularly suitable for real-time detection in resource-constrained environments, such as UAVs or embedded systems. This study provides an efficient and accurate solution for highland barley spike detection, contributing to growth monitoring, and precision agricultural management. Furthermore, the proposed lightweight architecture demonstrates potential for broader application in other agricultural contexts, offering a reference framework for real-time crop detection using UAV technology.

Acknowledgements

Appreciations are given to the editors and reviewer of the Journal Plant Method.

Author contributions

H.Z. and M.C. conceived the project. M.C., H.D., J.C., W.G., Z.H., D.Y. and H.Z. designed and conducted the experiments. H.Z. collected the data. M.C., H.D. and J.C. analyzed the data. H.Z. and M.C. wrote the manuscript with reviews from all the authors.

Funding

This study was supported by the Tibet Autonomous Region Science and Technology Plan Project Key Project (XZ202201ZY0003G), National Natural Science Foundation of China (32371853), and Natural Science Foundation of Fujian Province (2021J01059).

Data availability

The dataset can be available from https://modelscope.cn/datasets/Cai121/highland_barley and the code can be available from <https://github.com/trangle666ddd/YOLOv5-highland-barley-detection>.

Declarations

Competing interests

The authors declare that they have no Competing interests.

Received: 30 September 2024 Accepted: 27 February 2025

Published online: 24 March 2025

References

- Obadi M, Qi Y, Xu B. Highland barley starch (Qingke): structures, properties, modifications, and applications. *Int J Biol Macromol*. 2021;185:725–38. <https://doi.org/10.1016/j.jbiomac.2021.06.204>.
- Obadi M, Sun J, Xu B. Highland barley: chemical composition, bio-active compounds, health effects, and applications. *Food Res Int*. 2021;140:110065. <https://doi.org/10.1016/j.foodres.2020.110065>.
- Zhou C, Li M, Xiao R, Zhao F, Zhang F. Significant nutritional gaps in Tibetan adults living in agricultural counties along Yarlung Zangbo River. *Front Nut*. 2022;9:845026. <https://doi.org/10.3389/fnut.2022.845026>.
- Ren S, Chen H, Hou J, Zhao P, Feng H. Based on historical weather data to predict summer field-scale maize yield: Assimilation of remote sensing data to WOFOST model by ensemble Kalman filter algorithm. *Comput Elect Agric*. 2024;219:108822. <https://doi.org/10.1016/j.compag.2024.108822>.
- Jiang Y, Li C, Xu R, Sun S, Robertson J, Paterson A. DeepFlower: A deep learning-based approach to characterize flowering patterns of cotton plants in the field. *Plant Methods*. 2020;16:1–17. <https://doi.org/10.1186/s13007-020-00698-y>.
- Zhou Q, Huang Z, Zheng S, Jiao L, Wang L, Wang R, A wheat spike detection method based on transformer. *Front Plant Sci*. 2022;12:1023924. <https://doi.org/10.3389/fpls.2022.1023924>.
- Yang S, Hu L, Wu H, Ren H, Qiao H, Li P, Fan W. Integration of crop growth model and random forest for winter wheat yield estimation from UAV hyperspectral imagery. *IEEE J Selected Topics Appl Earth Obser Remote Sens*. 2021;14:6253–69. <https://doi.org/10.1109/JSTARS.2021.3089203>.
- Ye Z, Wei J, Lin Y, Guo Q, Zhang J, Zhang H, Deng H, Yang K. Extraction of olive crown based on UAV visible images and the U²-net deep learning model. *Remote Sens*. 2022;14(6):1523. <https://doi.org/10.3390/rs14061523>.
- Guo Q, Zhang J, Guo S, Ye Z, Deng H, Hou X, Zhang H. Urban tree classification based on object-oriented approach and random forest algorithm using unmanned aerial vehicle (UAV) multispectral imagery. *Remote Sens*. 2022;14(16):3885. <https://doi.org/10.3390/rs14163885>.
- Li J, Li J, Zhao X, Su X, Wu W. Lightweight detection networks for tea bud on complex agricultural environment via improved YOLO v4. *Comput Elect Agric*. 2023;211:107955. <https://doi.org/10.1016/j.compag.2023.107955>.
- Du X, Cheng H, Ma Z, Lu W, Wang M, Meng Z, Jiang C, Hong F. DSW-YOLO: A detection method for ground-planted strawberry fruits under different occlusion levels. *Comput Elect Agric*. 2023;214:108304. <https://doi.org/10.1016/j.compag.2023.108304>.
- Liu G, Hu Y, Chen Z, Guo J, Ni P. Lightweight object detection algorithm for robots with improved YOLOv5. *Eng Appl Artif Intell*. 2023;123:106217. <https://doi.org/10.1016/j.engappai.2023.106217>.
- Jia Y, Fu K, Lan H, Wang X, Su Z. Maize tassel detection with CA-YOLO for UAV images in complex field environments. *Comput Elect Agric*. 2024;217:108562. <https://doi.org/10.1016/j.compag.2023.108562>.

14. Yun J, Jiang D, Liu Y, Sun Y, Tao B, Kong J, Tian J, Tong X, Xu M, Fang Z. Real-time target detection method based on lightweight convolutional neural network. *Front Bioeng Biotechnol*. 2022;10:861286. <https://doi.org/10.3389/fbioe.2022.861286>.
15. Cui M, Lou Y, Ge Y, Wang K. LES-YOLO: A lightweight pinecone detection algorithm based on improved YOLOv4-Tiny network. *Comput Elect Agric*. 2023;205: 107613. <https://doi.org/10.1016/j.compag.2023.107613>.
16. Mendes PA, Coimbra AP, de Almeida AT. Forest vegetation detection using deep learning object detection models. *Forests*. 2023;14(9):1787. <https://doi.org/10.3390/f14091787>.
17. Yang H, Lin D, Zhang G, Zhang H, Wang J, Zhang S. Research on detection of rice pests and diseases based on improved yolov5 algorithm. *Appl Sci*. 2023;13(18):10188. <https://doi.org/10.3390/app131810188>.
18. Yao J, Qi J, Zhang J, Shao H, Yang J, Li X. A real-time detection algorithm for kiwifruit defects based on YOLOv5. *Electronics*. 2021;10(14):1711. <https://doi.org/10.3390/electronics10141711>.
19. Chen S, Liao Y, Lin F, Huang B. An improved lightweight YOLOv5 algorithm for detecting strawberry diseases. *IEEE Access*. 2023;11:54080–92. <https://doi.org/10.1109/ACCESS.2023.3282309>.
20. Yu G, Cai R, Luo Y, Hou M, Deng R. A-pruning: A lightweight pineapple flower counting network based on filter pruning. *Complex Intell Syst*. 2024;10(2):2047–66. <https://doi.org/10.1007/s40747-023-01261-7>.
21. Kong C, Yang L, Gong H, Wang L, Li H, Li Y, Wei B, Nima C, Deji Y, Zhao S, et al. Dietary and food consumption patterns and their associated factors in the Tibetan Plateau population: Results from 73 counties with agriculture and animal husbandry in Tibet, China. *Nutrients*. 2022;14(9):1955. <https://doi.org/10.3390/nu14091955>.
22. Chun W. Make the golden highland barley fuller: Shigatse City actively explores a new path for the high-quality development of the highland barley industry (in Chinese). (2023). http://nynct.xizang.gov.cn/xwzx/ztzl/gdzt/ztzg/202311/t20231114_388127.html. Accessed 23 Dec 2023.
23. Ye Z, Yang K, Lin Y, Guo S, Sun Y, Chen X, Lai R, Zhang H. A comparison between pixel-based deep learning and object-based image analysis (OBIA) for individual detection of cabbage plants based on UAV visible-light images. *Comput Elect Agric*. 2023;209:107822. <https://doi.org/10.1016/j.compag.2023.107822>.
24. Liu B, Su S, Wei J. The effect of data augmentation methods on pedestrian object detection. *Electronics*. 2022;11(19):3185. <https://doi.org/10.3390/electronics11193185>.
25. Chollet F. Xception: Deep learning with depthwise separable convolutions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017. p. 1251–1258. <https://doi.org/10.1109/CVPR.2017.195>.
26. Fang Z, Ren J, Marshall S, Zhao H, Wang S, Li X. Topological optimization of the DenseNet with pretrained-weights inheritance and genetic channel selection. *Pattern Recogn*. 2021;109:107608. <https://doi.org/10.1016/j.patcog.2020.107608>.
27. Cao J, Bao W, Shang H, Yuan M, Cheng Q. GCL-YOLO: a GhostConv-based lightweight yolo network for UAV small object detection. *Remote Sens*. 2023;15(20):4932. <https://doi.org/10.3390/rs15204932>.
28. Ye Z, Guo Q, Wei J, Zhang J, Zhang H, Bian L, Guo S, Zheng X, Cao S. Recognition of terminal buds of densely-planted Chinese fir seedlings using improved YOLOv5 by integrating attention mechanism. *Front Plant Sci*. 2022;13: 991929. <https://doi.org/10.3389/fpls.2022.991929>.
29. Woo S, Park J, Lee J, Kweon I. CBAM: Convolutional block attention module. In: *Proceedings of the European Conference on Computer Vision*. 2018. p. 3–19. https://doi.org/10.1007/978-3-030-01234-2_1.
30. Zhang J, Chen H, Yan X, Zhou K, Zhang J, Zhang Y, Jiang H, Shao B. An improved YOLOv5 underwater detector based on an attention mechanism and multi-branch reparameterization module. *Electronics*. 2023;12(12):2597. <https://doi.org/10.3390/electronics12122597>.
31. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen L. Mobilenetv2: Inverted residuals and linear bottlenecks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018. p. 4510–4520. <https://doi.org/10.1109/CVPR.2018.00474>.
32. Howard A, Sandler M, Chu G, Chen LC, Chen B, Tan M, Wang W, Zhu Y, Pang R, Vasudevan V, et al. Searching for mobilenetv3. In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2019. p. 1314–1324. <https://doi.org/10.1109/ICCV.2019.00140>.
33. Ma N, Zhang X, Zheng HT, Sun J. Shufflenet v2: Practical guidelines for efficient cnn architecture design. In: *Proceedings of the European conference on computer vision (ECCV)*. 2018. p. 116–131. https://doi.org/10.1007/978-3-030-01264-9_8.
34. Soviany P, Ionescu RT. Optimizing the trade-off between single-stage and two-stage object detectors using image difficulty prediction. In: *2018 20th international symposium on symbolic and numeric algorithms for scientific computing (SYNASC 2018)*. IEEE; 2019. p. 209–14. <https://doi.org/10.1109/SYNASC.2018.00041>.
35. Chen R, Chen Y. Improved convolutional neural network YOLOv5 for underwater target detection based on autonomous underwater helicopter. *J Marine Sci Eng*. 2023;11(5):989. <https://doi.org/10.3390/jmse11050989>.
36. Wang C, Bochkovskiy A, Liao H. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023. p. 7464–7475. <https://doi.org/10.1109/CVPR52729.2023.00721>.
37. Jocher G, Chaurasia A, Qiu J. Ultralytics YOLO. 2023. <https://github.com/ultralytics/ultralytics>. Access 12 June 2023.
38. Yu Z, Lei Y, Shen F, Zhou S, Yuan Y. Research on identification and detection of transmission line insulator defects based on a lightweight YOLOv5 network. *Remote Sens*. 2023;15(18):4552. <https://doi.org/10.3390/rs15184552>.
39. Zhang B, Wang R, Zhang H, Yin C, Xia Y, Fu M, Fu W. Dragon fruit detection in natural orchard environment by integrating lightweight network and attention mechanism. *Front Plant Sci*. 2022;13:1040923. <https://doi.org/10.3389/fpls.2022.1040923>.
40. Sun Z, Yang H, Zhang Z, Liu J, Zhang X. An improved YOLOv5-based tapping trajectory detection method for natural rubber trees. *Agriculture*. 2022;12(9):1309. <https://doi.org/10.3390/agriculture12091309>.
41. Yang G, Wang J, Nie Z, Yang H, Yu S. A lightweight YOLOv8 tomato detection algorithm combining feature enhancement and attention. *Agronomy*. 2023;13(7):1824. <https://doi.org/10.3390/agronomy13071824>.
42. Chen Z, Chen J, Li Y, Gui Z, Yu T. Tea bud detection and 3D pose estimation in the field with a depth camera based on improved YOLOv5 and the optimal pose-vertices search method. *Agriculture*. 2023;13(7):1405. <https://doi.org/10.3390/agriculture13071405>.
43. Zhang X, Min C, Luo J, Li Z. YOLOv5-FF: Detecting floating objects on the surface of fresh water environments. *Appl Sci*. 2023;13(13):7367. <https://doi.org/10.3390/app13137367>.
44. Guo S, Yoon S, Li L, Wang W, Zhuang H, Wei C, Liu Y, Li Y. Recognition and positioning of fresh tea buds using YOLOv4-lighted+ICBAM model and RGB-D sensing. *Agriculture*. 2023;13(3):518. <https://doi.org/10.3390/agriculture13030518>.
45. Dvornik N, Mairal J, Schmid C. On the importance of visual context for data augmentation in scene understanding. *IEEE Trans Pattern Anal Machine Intell*. 2019;43(6):2014–28. <https://doi.org/10.1109/TPAMI.2019.2961896>.
46. Du J, Li J, Fan J, Gu S, Guo X, Zhao C. Detection and identification of tassel states at different maize tasseling stages using UAV imagery and deep learning. *Plant Phenomics*. 2024;6:0188. <https://doi.org/10.34133/plantphenomics.0188>.
47. Russell B, Torralba A, Murphy K, Freeman W. LabelMe: a database and web-based tool for image annotation. *Int J Comput Vision*. 2008;77:157–73. <https://doi.org/10.1007/s11263-007-0090-8>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.