



Original research article

Using UAVRS and deep learning to conduct resource surveys of threatened Tibetan medicinal plants in the Qinghai-Tibet Plateau



Chenghui Wang^{a,b}, Ziyi Li^b, Rong Ding^c, Jiawei Luo^d, Yu Liang^b, Rui Gu^{b,e,*}, Shihong Zhong^{a,**}

^a School of Pharmacy, Southwest Minzu University, Chengdu 610041, China

^b School of Ethnic Medicine, Chengdu University of Traditional Chinese Medicine, Chengdu 611137, China

^c School of Pharmacy, Chengdu University of Traditional Chinese Medicine, Chengdu 611137, China

^d West China Biomedical Big Data Center, West China Hospital/West China School of Medicine, Sichuan University, Chengdu 610044, China

^e State Key Laboratory of Southwestern Chinese Medicine Resources, School of Pharmacy, Chengdu University of Traditional Chinese Medicine, Chengdu 611137, China

ARTICLE INFO

Keywords:

Gentiana szechuenensis
Gentiana veitchiorum
Unmanned aerial vehicle remote sensing (UAVRS)
Deep learning
Medicinal plant

ABSTRACT

Gentiana szechuenensis Kanitz. and *Gentiana veitchiorum* Hemsl. are two wild medicinal plants widely used in Tibetan medicine. In recent decades, their wild populations have declined rapidly due to persistent over-harvesting of their flowers. Because their flowers are small and dense, it is difficult to rely on manual counting under wild conditions. The combination of deep learning and unmanned aerial vehicle remote sensing (UAVRS) is a new method for plant surveys. We trained 12 of the most advanced or widely used YOLO models on a custom dataset to achieve quantitative detection of both plant flowers in UAVRS images. The accuracy, precision and recall of *G. szechuenensis* Flower (GSF) detection can reach 97.00%, 90.40% and 95.40%, respectively (based on YOLOv7). Similarly, those of *G. veitchiorum* Flower (GVF) detection can reach 93.40%, 91.30% and 90.60%, respectively (based on YOLOv5n), and the highest mAP can reach 94.90% (based on YOLOv5m). Based on the detection results, it is calculated that the total yield of dried GSF and GVF that can be harvested as medicinal materials in study area ranges from 1.88 to 2.10 g·m⁻² and 4.42–4.62 g·m⁻², respectively. The results show that deep learning and UAVRS can be used for quantitative detection of GSF and GVF, which is helpful for further research and protection of these two plants.

1. Introduction

The investigation of wild plant resources is crucial for plant protection and ecosystem management (Corlett, 2016). Providing accurate, timely, and detailed investigation results can greatly help to avoid loss of biodiversity and over-harvesting (Hunault-Fontbonne and Eydinson, 2023). However, wild plants often receive less attention compared to some widely known animals, leading to a more severe situation of resource scarcity in plant protection than animal protection (Goetsch et al., 2015). As a result, investigators of plant resources often face more severe challenges in terms of accuracy, efficiency, cost, labor protection, and how to reduce

* Corresponding author at: School of Ethnic Medicine, Chengdu University of Traditional Chinese Medicine, Chengdu 611137, China.

** Corresponding author.

E-mail addresses: gurui@cdutcm.edu.cn (R. Gu), 527455247@qq.com (S. Zhong).

interference with wild populations (Chi et al., 2017; Corlett, 2023).

There are mainly two methods for surveying wild plants: ground recording and remote sensing monitoring. Ground recording usually adopts sampling methods, such as quadrat method, belt transect method, line transect method, etc. (Pohjanmies et al., 2021). For example, based on 11 quadrats of 25 m×25 m for different vegetation types, quantitative data on plant species richness and abundance were obtained for the island of St. Eustatius in the Caribbean Netherlands (van Andel et al., 2016). Similarly, Nunes et al. (2022) conducted a quantitative analysis of communities with different levels of protection using belt transects and quadrats. The advantage of ground recording is that it can directly observe and collect information on plant populations. However, ground recording also has some drawbacks that are difficult to overcome, such as low efficiency, subjectivity, high cost, and terrain limitations (Peng et al., 2020). Remote sensing monitoring, through aerial or space observation platforms, has completed a new interpretation of the plant resource survey task, far exceeding the traditional sampling survey in terms of survey scope and flexibility, and also plays an increasingly important role in plant surveys, such as mapping phytoplankton in river systems (VonBank et al., 2017) and seagrass in coastal ecosystems (Coffer et al., 2023). Although remote sensing monitoring has a high degree of flexibility in terms of survey time, area and range, and the non-contact survey basically does not affect the original plant community (Beamish et al., 2020; Matas-Granados et al., 2022), it also has some limitations. For example, flying manned aircraft to conduct plant surveys in areas with complex terrain is risky (Dash et al., 2019), and it is usually relatively expensive for plant survey tasks that lack financial support, while satellite remote sensing is difficult to meet the needs of satisfactory plant community surveys even with very high-resolution satellite images (≤ 1 m) (Royimani et al., 2019; Zhang et al., 2023), and is more often used to perform large-scale vegetation surveys, such as vegetation type classification (Trujillo-Jiménez et al., 2022), vegetation coverage estimation (Zhang et al., 2022) and invasive plant mapping (Theron et al., 2022).

As the level of hardware manufacturing improves and the cost of electronic integrated devices decreases, unmanned aerial vehicle remote sensing (UAVRS) is rapidly developing (Colomina and Molina, 2014; Jiménez López and Mulero-Pázmány, 2019). Compared to traditional photogrammetry methods such as ground surveys, UAVRS can obtain continuous image coverage of tens of hectares in one flight (Kupková et al. 2023). In contrast to satellite remote sensing imagery, UAVRS images can achieve millimeter-scale spatial resolution (Nakamura et al., 2017). Due to its significant price advantage, UAVRS has replaced most remote sensing data collection work that was previously performed by traditional manned aircraft (Puri et al., 2017). When facing survey subjects such as *Gentiana szechenyi* flower (GSF) and *Gentiana veitchiorum* flower (GVF), which have a corolla size of less than 2 cm, UAVRS can provide both millimeter-scale spatial resolution and hectare-scale image coverage simultaneously, thus maximizing the structural features of the flowers while maintaining large areas of coverage. With excellent cost, efficiency, accuracy, and safety advantages, UAVRS is increasingly used in wild plant surveys (Mao et al., 2023; Scheeres et al., 2023; Valente et al., 2022).

However, excellent spatial resolution and coverage can also become a burden for researchers, as each UAVRS survey can produce a large aerial image dataset covering a large area. If staff manually identify each image, it will be tedious. For example, in our study, over 270,000 blooms can be collected in a hectare of land. It is obviously impractical for humans to count so many flowers manually. To address this issue, researchers have developed semi-automatic techniques that combine machine learning to detect plants and specific organs in UAVRS images (da Silva et al., 2023; Fu et al., 2022; Hamylton et al., 2020; Wan Nurazwin Syazwani et al., 2022). The simplest and most commonly used methods are pixel-based classification methods such as threshold segmentation, supervised classification, and unsupervised clustering (Mafanya et al., 2017). These methods achieve good recognition accuracy by extracting and quantifying the spectral differences between target and background values. However, their performance is often poor in complex environments (Sotille et al., 2022). To improve recognition accuracy in such environments, researchers often introduce more complex sensors, such as multispectral, hyperspectral or even lidar (Fu et al., 2023; Ishida et al., 2018; Wolff et al., 2023), and combine complex data preprocessing steps to increase the spectral differences between different objects and achieve accurate recognition results (Adar et al., 2022; Fernández-Guisuraga et al., 2022; Zhao et al., 2020). For instance, Su et al. (2022) used multispectral images obtained by UAVRS, combined with feature selection algorithms and machine learning algorithms, for blackgrass detection in images, achieving better results than using all features. Although non-deep learning-based machine learning methods have achieved encouraging detection results in plant surveys by enhancing spectral differences, these methods increase the cost of the survey and cannot achieve complete automation due to the complexity of the process and the need for continuous human participation. Furthermore, they cannot fully learn complex plant individual characteristics such as structure and shape.

As one of the important research directions of machine learning, deep learning-based object detection models using convolutional neural networks have profoundly changed the way vegetation identification and plant individual detection are performed (Ding et al., 2023; Oh et al., 2020). Based on supervised learning strategies, these models can effectively leverage image-level annotated datasets to automatically learn relevant features of objects while training (Zhang et al., 2020), which is crucial for assigning correct categories to observed targets in images. High-performance object detection algorithms coupled with UAVRS images with millimeter-scale spatial resolution provide the possibility of replacing time-consuming and labor-intensive field resource investigation scenarios (Zhang et al., 2021). This has the potential to change the traditional mode of wild plant resource investigation, achieving large-scale, high-precision, and low-cost deployment and continuous monitoring.

The execution process of existing advanced object detection models mainly adopts two strategies: (1) The two-stage detection strategy first generates a region proposal anchor box for all possible regions in the input image that may be targets, and then performs classification and regression tasks on the region proposal anchor box to complete the entire target detection process; (2) The one-stage detection does not require the generation of region proposal anchors in advance, but instead performs regression tasks on the position of the target bounding box and the category probability of the bounding box in the input image simultaneously. Object detection models using a one-stage detection strategy usually have faster detection speeds and lower requirements for computing resources. Therefore, compared with region proposal-based two-stage models such as Fast R-CNN (Girshick, 2015), Faster R-CNN (Ren et al.,

2015), and Mask R-CNN (He et al., 2017), they are more suitable for deployment in specific practical application scenarios. For example, the YOLO series, the most famous state-of-the-art (SOTA) algorithm, has been iteratively formed by different versions such as YOLO (Redmon et al., 2016), YOLOv3 (Redmon and Farhadi, 2018), YOLOv4 (Bochkovskiy et al., 2020), YOLO-R (Wang et al., 2021), YOLOv5 (Jocher et al., 2021), YOLOv6 (Li et al., 2022), YOLOv7 (Wang et al., 2023), YOLOv8 (Jocher et al., 2023) and their corresponding derivative versions. Though researchers widely use the YOLO series for plant object detection and counting tasks (Bomantara et al., 2023; Nan et al., 2023) and achieved great accuracy. The applicability of these efficient models to dense plant surveys in real complex scenarios, such as the GSF and GVF detection in this study, remains unclear. Moreover, under extensive field conditions, surveyors can usually only carry laptops with limited performance to perform survey tasks, which prevents them from using powerful GPUs to speed up the detection process. Therefore, the feasibility of using edge performance computing devices to apply deep learning to plant surveys in UAVRS images deserves further study.

Given the above issues, the main objective of this study is to establish a realistic plant survey scenario using deep learning and UAVRS to count the number of flowers of *G. szechenyii* and *G. veitchiorum*, in the Qinghai-Tibet Plateau region and calculate the harvestable yield for medicinal use. Additionally, the study compares the application capabilities of recent YOLO series models under performance limitations and explores the feasibility of using UAVRS and deep learning for wild plant resource surveys. We hope that this study can demonstrate more intuitively that UAVRS and deep learning provide a favorable choice for investigating wild plant resources, encouraging more researchers in endangered plant protection to consider using UAVRS and artificial intelligence to survey other plants in ecologically fragile areas and select appropriate models based on our research findings.

2. Materials and methods

2.1. Study Area and target plants

The study site is located in Zeku County, Qinghai Province, China. Zeku County is located within the Sanjiangyuan Nature Reserve and has a high-altitude continental climate characterized by simultaneous heat and rain, abundant and concentrated precipitation, and is highly suitable for the growth of forage grasses. The unique geographical location conditions make Zeku County an important water source protection area in China and one of the most important animal husbandry bases on the Qinghai-Tibet Plateau. Alpine meadows are the main land type in Zeku County, accounting for more than 96% of the county's area. These alpine meadows are home to many wild populations of Tibetan medicinal materials commonly used by Tibetans and are also the main distribution areas for *G. szechenyii* and *G. veitchiorum*. It is an ideal research site for image collection of two gentian species. The specific research site is a flat grassland with a large number of *G. szechenyii* and *G. veitchiorum* distributed. The coordinates of the center point are: 35°4'35.95"N, 101°33'31.52"E, with an altitude of 3753.13 m (Fig. 1).

These two medicinal plants in this study belong to the family Gentianaceae, which *G. szechenyii* Kanitz. belongs to section *Isomeria* Kusn, while *G. veitchiorum* Hemsl. belongs to section *Kudoa* (Masamune) Satake & Toyokuni ex Toyokuni (Favre et al., 2020). Both plants are perennial herbs, with dense clusters of medium-sized flowers, slightly more in number, short sepals, only one-third to one-half of the corolla length, and lobes usually shorter than the sepal tube. The corolla is funnel-shaped, with a length of 4–6 cm. However, *G. szechenyii* Kanitz has a very short stem and the basal leaves are densely stacked into a lotus-like shape, and the upper part of the corolla is white, the lower part is yellow-white, with blue-gray spots or wide stripes. *G. veitchiorum* Hemsl. has several pairs of stem leaves, which become denser and longer towards the upper part of the stem, and the upper part of the corolla is dark blue, the lower part is yellow-green, and has dark blue spots and stripes. The apparent difference in flower color is reflected in the names of the two plants' dried flowers used as medicine by the Tibetan people. The dried flowers of *G. szechenyii* Kanitz are called Bangjian-Gabao (羌珍嘎波), meaning white-flowered gentians, while the dried flowers of *G. veitchiorum* Hemsl. are called Bangjian-Wenbao (羌珍文波), meaning blue-flowered gentians (Li et al., 2023).

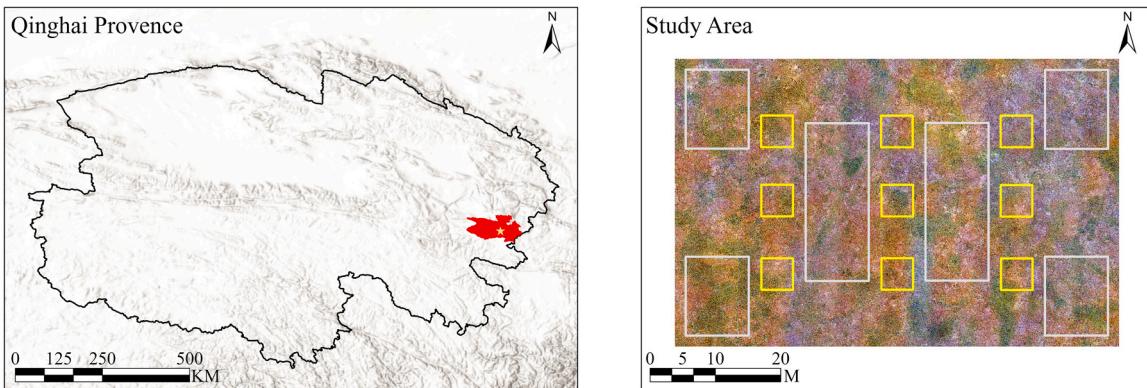


Fig. 1. The location of the study area, shown in the figure. The study area is located in the high-altitude meadows of Zeku County, Qinghai Province. In the orthomosaic obtained by stitching UAVRS images, the plots used to construct the custom dataset are shown in white outline, and the plots used to compare the consistency of the object detection model with professionals are shown in yellow outline.

Both the two medicines are widely used in traditional Tibetan cuisine and for treating respiratory diseases such as pneumonia, cough, tracheitis, and laryngitis (Chi et al., 2021; Zhang et al., 2014). The GSF and GVF that can be bought on the market are all derived from wild plant resources, among which GSF is an endangered Tibetan medicine of the first class, and GVF is a substitute for GSF. During the COVID-19 epidemic, GSF and GVF were collected in copious quantities and used to prepare Tibetan medicines to treat respiratory system diseases, further increasing the pressure on the wild resources. Therefore, strengthening the investigation, evaluation, and protection of the wild resources of GSF and GVF is of great significance for the sustainable development of the Tibetan medicine industry. The wild populations of *G. szechuenii* Kanitz. and *G. veitchiorum* Hemsl. are distributed in the alpine meadows of the Qinghai-Tibet Plateau at 3000–4800 m. The former is mostly patchy in suitable grasslands, while the latter can cover a large area in the meadows. The two species have similar habitats and can coexist in the same meadow community (Fig. 2).

2.2. UAVRS data acquisition and post-processing

Both the two plants have a long flowering period, with *G. szechuenii* flowering from June to October and *G. veitchiorum* flowering from June to October (Committee, 2018; Ho and Liu, 2001). In September, both plants enter their peak flowering period, which is also the traditional time for harvesting as medicinal materials. Therefore, in this study, UAVRS images were obtained on September 11, 2021, which was a clear, calm day, and the exact collection time was 14:05–14:34 local time. The images of the study area were obtained using a DJI Mavic 2 Pro UAVRS (DJI Technology Co., Ltd., Shenzhen, China), which is equipped with a set of GNSS satellite positioning system and has a built-in 20-megapixel Hasselblad camera (i.e., 5472 by 3648 pixels per image). Pix4D capture software (Pix4D, S.A., Lausanne, Switzerland) was used to set flight plan parameters and control the UAVRS to take pictures according to the set shooting program during flight. We found that due to occasional gusts of wind in the Qinghai-Tibet Plateau region, not only does it affect the stability of UAVRS flight, but it also causes GSF and GVF to sway. In windy conditions, if aerial photography using an equidistant interval shooting strategy is used, due to the high relative motion speed between the camera and the flowers being photographed, it will result in blurred images. Therefore, we used waypoint hovering photography combined with manual focusing to obtain images to ensure shooting stability as much as possible. The photo overlap rate along the flight direction was 70%, and the photo overlap rate along the lateral direction was 60%. The flight speed was $6.8 \text{ m}\cdot\text{s}^{-1}$ and the flight time was 30 minutes. The average flight altitude was 10 m. During data collecting, the camera was set to auto-white balance mode and no color calibration or ground control was used. Due to a certain slope in the study area, the takeoff altitude was based on the distance from the center point of the study area. The collected 499 UAVRS images were orthorectified according to Agisoft Metashape Pro (Agisoft LLC, St. Petersburg, Russia) standard motion structure SfM workflow. After cropping out parts with insufficient edge quality in the software and converting the coordinate system to World Geodetic System 84-Universal Transverse Mercator coordinate system Zone 48 N (WGS 84/UTM zone 48 N), an orthomosaic was output at a ground sampling distance (GSD) of 0.23 cm for images. The resulting image resolution was 36,

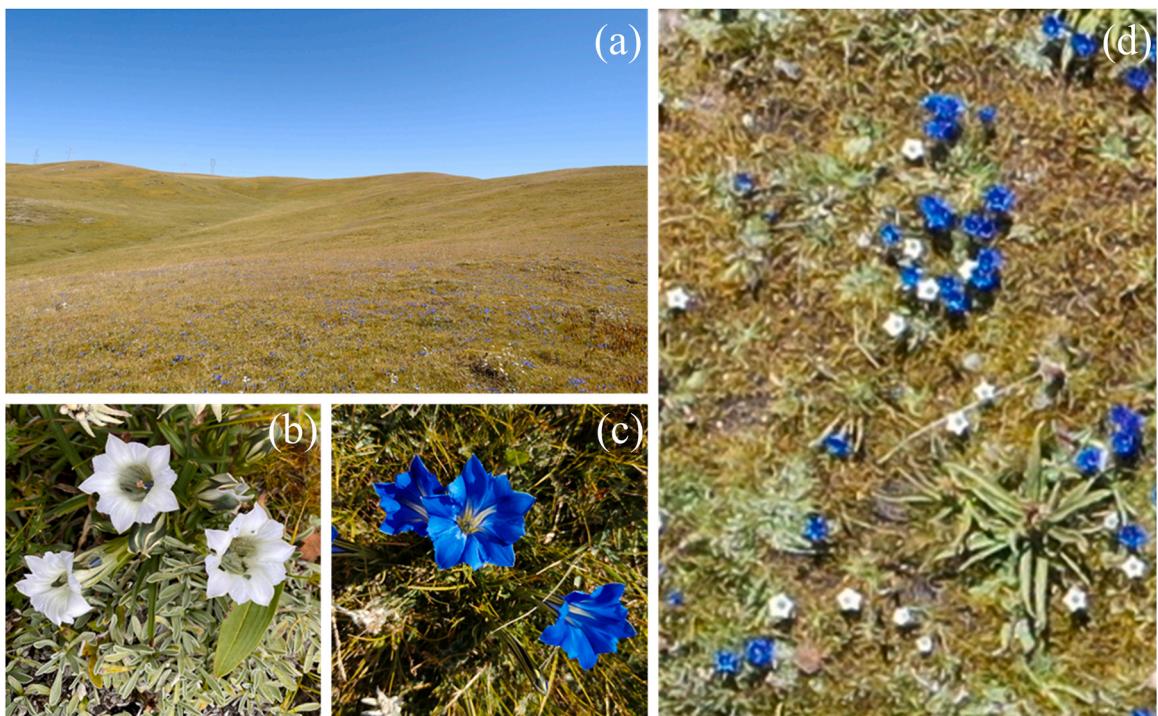


Fig. 2. Typical high-density *G. szechuenii* and *G. veitchiorum* habitat (a), *G. szechuenii* in bloom (corolla diameter ~2 cm) (b), *G. veitchiorum* in bloom (corolla diameter ~2 cm) (c), *G. szechuenii* and *G. veitchiorum* blossoms as seen from 10-m altitude UAVRS imagery (d).

084×23,348 pixels and covered an area of 2957.13 m².

2.3. Datasets preparation

In this study, the image dataset is divided into two categories. The first category is a dataset used to train, validate, and test deep learning-based object detection models. First, 2 mosaics of 12,800×5120 pixels and 4 mosaics of 6400×5120 pixels were cropped from the orthomosaic images in the study area, and then divided into separate jpg images, each image being 512×512 pixels, for a total of 1000 images. The second category is a dataset for comparing the consistency of object detection models and professionals in detecting two types of flowers. First, 9 mosaics of 2560×2560 pixels were cropped and then divided into a jpg image set of 512×512 pixels, for a total of 225 images. The cropping of mosaics was done by using ENVI5.6 (Exelis Visual Information Solutions, Boulder, CO, USA) software.

This study used the Python-based data annotation tool LabelImg to annotate the dataset (Chen et al., 2013). The specific data annotation process was for two professionals with experience in identifying GSF and GVF to manually draw rectangular bounding boxes around the corolla of the flowers and record the flower category. Then an additional staff member manually corrected it, including correcting possible wrong selections and missed flowers to ensure the reliability of the reference data. We annotated 1000 images with a total of 33,091 annotations, including 8989 GSF annotations and 24,102 GVF annotations. The annotated images were stored in the PASCAL VOC dataset format and divided into training, validation, and test sets at a ratio of 8:1:1. Data augmentation was performed separately according to the dataset category. Cropping, translation, rotation angle, mirroring, changing brightness, cutout, adding noise and other means were selected as data augmentation methods. By augmenting the initial dataset, various objects and backgrounds can be enriched, thereby reducing overfitting and improving the generalization ability of the model (Kamilaris and Prenafeta-Boldú 2018). Finally, 33,091 annotated samples and their corresponding ground images were increased to 165,455 samples. This method can make the trained model more robust when facing changes in light and viewpoint as well as other minor changes caused by different outdoor conditions or equipment.

2.4. Data processing

2.4.1. Object detection model

Because of the great success that the “You Only Look Once” series have made in object detection, we compared the main models of the recent 6 versions of the YOLO series for the detection of GSF and GVF in the dataset. Considering that YOLOv5 is one of the most used object detection algorithms, while YOLOv7 and YOLOv8 are the latest algorithms, their smaller and larger models may better reflect the performance range in real scenarios, so they were also included in the study scope. Therefore, the models examined in this study were YOLOv3, YOLOv4, YOLO-R, YOLOv5n, YOLOv5m, YOLOv5x, YOLOv7tiny, YOLOv7, YOLOv7x, YOLOv8n, YOLOv8m and YOLOv8x. It was built using the Python-based PyTorch deep learning development framework.

Generally speaking, although the model’s performance can be achieved by increasing the network scale, it requires processing time as a cost. That is to say, larger models can achieve higher detection accuracy when facing complex large-scale detection tasks but require more processing time than smaller models. In this study, we used and compared the performance of the 12 YOLO models in two plant flower detection tasks, focusing on detection accuracy and detection efficiency.

Based on the superior performance and low memory requirements of the YOLO series, we trained all models for 300 epochs with a batch size (i.e., the number of image samples passed to the detector at the same time) of 32. During training, we used the model weights with the highest mAP on the validation set to test each model. Depending on the complexity of the model, training took about 1.7 hours to 14.9 hours. We retained the model weights that achieved mAP on the validation set for subsequent image inference.

The experimental platform is equipped with an Intel® Core™ i9-10900X CPU @ 3.70 GHz, an NVIDIA GeForce RTX 3090 GPU with 24 GB of video memory and runs on a 64-bit Windows 10 operating system. The software environments include CUDA 12.0, CUDNN 8.1.1, and Python 3.8 as the compiler.

2.4.2. Performance evaluation

In this study, average precision (AP) and mean average precision (mAP) were used to evaluate the accuracy of different YOLO models in GSF and GVF detection, and their corresponding calculation methods are as follows:

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

$$mAP = \frac{1}{n} \sum_{i=1}^n AP_i \quad (2)$$

Where n represents the total number of categories, which is 2 (GSF and GVF) in this study. P and R represent Precision and Recall, respectively, and are calculated as follows:

$$Precision = \frac{TP}{(TP + FP)} \quad (3)$$

$$\text{Recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \quad (4)$$

Where TP is an abbreviation for true positive, representing the number of predicted anchor boxes with $\text{IoU} > 0.5$, FP is an abbreviation for false positive, representing the number of predicted anchor boxes with $\text{IoU} \leq 0.5$, and FN is an abbreviation for false negative, representing the number of labels that were not successfully predicted.

After training, the training time, model weight size, and GFLOPs data of each model were recorded as efficiency indicators for measuring the complexity of the model. In order to better evaluate the ability of different models in balancing accuracy and available computational resources, the efficiency and accuracy indicators were normalized. Considering that the evaluation standard of the accuracy indicator is the larger the better, while the efficiency indicator is the opposite, the normalized efficiency indicator was reversed processed, that is, 1 was subtracted from the normalized efficiency indicators. Finally, the indicators of each model were summed up and sorted according to the scoring situation. The higher the score of the model, the better the balance of its performance and efficiency.

After preliminarily evaluating the performance of each YOLO model, we used 5-fold cross-validation to verify the stability of the models involved in this study. In this method, we merged the training set and validation set used for target detection model training, and randomly divided them into 5 subsets. We then used 4 of these subsets as the training set for model training, and the remaining one subset as the validation set for model evaluation, repeating this process 5 times. The data were compared using a one-way ANOVA and post hoc Tukey tests. P values of <0.05 were considered statistically significant.

2.4.3. Consistency evaluation of object detection model and professionals in detection

To verify whether the number of flowers detected by the trained object detection models corresponds to the actual number of flowers, two experienced people with GSF and GVF identification expertise counted the number of flowers in the 225 images of the consistency detection dataset, and then used all 12 models to perform flower detection on the dataset. The coefficient of determination (R^2) was used as the evaluation index to quantitatively evaluate the accuracy of different models on the number detection of GSF and GVF. The calculation formula of the coefficient of determination is as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (5)$$

Where n represents the total number of samples, y_i and \hat{y}_i represent the manually counted and model predicted values of the number of GSF or GVF in a single image, respectively, and \bar{y}_i is the average value of the number of GSF or GVF manually counted in all images.

2.4.4. Yield estimation - comparison of the performance of different CPUs in real survey scenarios

In the case of the Qinghai-Tibet Plateau region where the study area is located, for example, it is usually far from stable power and network supplies, which means that it is not possible to stably use more powerful computing devices such as desktop computers, gaming laptops or cloud computing tools because they rely on strong power supplies or stable network connections. Since ultrabooks generally carry CPUs set to lower or standard voltages, they have longer battery life. They are more common computing devices under working conditions like this, but at the same time, this means lower detection speed. In order to simulate the performance of different models in detecting tasks on ultrabooks under real field survey scenarios, we chose a typical ultrabook equipped with I5 10210 U (Lenovo ThinkPad X13 Gen1) for performance comparison, and selected two personal computers with more robust CPU as references (Table 1). All three devices were based on Python 3.8 to build separate virtual environments and configure the basic packages required for model operation.

After cropping the orthomosaic image of the entire study area into 512×512 datasets, GSF and GVF detection were performed using 12 YOLO models on three computing devices. The number of GSF and GVF in each image was counted based on the detection results, and the sum was used for yield estimation. The yield was calculated by constructing a relationship between flower weight and quantity. Specifically, after completing the UAVRS image acquisition work in the research area, 600 GSF and GVF flowers were collected in 6 batches according to the harvesting standards of medicinal materials. The flowers were brought back to the laboratory to dry and their weights under different quantities were measured. A linear regression equation was used to fit the relationship between the number and weight of the two kinds of flowers. The coefficient of determination was calculated to evaluate the rationality of

Table 1
The hardware and software conditions of the different devices involved.

PC	PC1	PC2	Laptop1
CPU	I9 10900X	I5 6600 T	I5 10210 U
Total Cores	10	4	4
Total Threads	20	4	8
Base Frequency	3.70 GHz	2.70 GHz	1.60 GHz
Max Turbo Frequency	4.50 GHz	3.50 GHz	4.20 GHz
Thermal Design Power	165 W	35 W	15 W
Memory	96 GB	16 GB	8 GB
Operating System	Windows 10	Windows 10	Windows 11

resource evaluation based on quantity.

3. Results

3.1. Detection results and analysis of different algorithms on the validation set

According to [Table 2](#), from the aspect of mean average precision (mAP), YOLOv5x achieved the best result of 95.00%, followed by YOLOv5m, YOLOv8n, and YOLOv5n with 94.90%, 94.90%, and 94.70%, respectively. Of the remaining 8 models, all except YOLOv7tiny and YOLOv4, which achieved only 90.50% and 92.80%, were above 94.00%. When discussing the average precision (AP) of GSF and GVF separately, it can be seen that poor GVF AP is an important factor in reducing the model's mAP, as the AP of GSF for the remaining 7 models were all above 96.00%, except for YOLOv7tiny, YOLOv4, YOLOv8m and YOLOv8x, which were only 92.00%, 94.80%, 94.90% and 95.60%. The top 4 scores in GVF detection were YOLOv5n, YOLOv8n, YOLOv5x, and YOLOv5m, with APs of 93.40%, 93.30%, 93.20%, and 93.10%, respectively, while the APs of the remaining 5 models did not exceed 93.00%.

The time it took for the 12 models to complete training varied. It only took 1.45 hours for YOLOv8n to complete the training, followed by YOLOv5n, YOLOv8m and YOLOv5m, which took 2.61 hours, 2.94 hours, and 3.30 hours, while the remaining 8 models, all took more than 5 hours to complete the whole training process. The weights obtained from training were arranged in ascending order as follows: YOLOv5n, YOLOv8n, YOLOv7tiny, YOLOv5m, YOLOv8m, YOLOv7, YOLO-R, YOLOv4, YOLOv3, YOLOv7x, YOLOv8x, and YOLOv5x, which was consistent with the order of GFLOPs commonly used to describe model complexity.

The results of the comprehensive ranking of accuracy and efficiency indicators show that, except for YOLOv7tiny, YOLOv4, and YOLOv5x, the total score ranking of the remaining 9 models is consistent with the Model size and GFLOPs indicators. Among them, YOLOv7tiny and YOLOv4 are greatly affected by the deficient performance in detection accuracy, which greatly affected the final ranking. Although YOLOv5x performs worst in terms of model size and GFLOPs, its good accuracy and training performance help it improve its ranking. YOLOv5n and YOLOv8n rank first and second respectively, due to their good accuracy and impressive volume advantages. The following are the intermediate scale models of YOLOv5, YOLOv8, and YOLOv7, indicating that newer models can better balance the performance of accuracy and efficiency.

To further measure the robustness of different models, 5-fold cross-validation was performed based on different datasets ([Fig. 3](#)). The results show that under the condition of training with different datasets, the five experimental results all show a high degree of consistency, and the standard deviation is all below 0.5%, indicating that after mixing, different training sets have similar information richness. It is worth noting that compared with the accuracy results in [Table 2](#), the average mAP, AP-GSF and AP-GVF of YOLOv7tiny have increased by 3.50%, 3.58% and 3.30%, respectively, while the fluctuation of accuracy indicators of other models is all below 1.00%, suggesting that YOLOv7tiny needs a larger dataset to achieve comparable detection accuracy with other models.

According to the example comparison of GSF and GVF detection results using different models ([Fig. 4](#)). The detection results of GSF and GVF are marked with rectangles of different colors. YOLOv3, YOLO-R, YOLOv7tiny, YOLOv7, YOLOv7x, YOLOv8m and YOLOv8x accurately identified each GSF in the image, while YOLOv4, YOLOv5n, YOLOv5m, YOLOv5x and YOLOv8n produced 1, 2, 1, 1 and 1 false positive results, respectively. In the detection of GVF, the detection results of YOLOv5n, YOLOv7tiny, and YOLOv8n are interpolated with the true values at 4, 3, and 5, respectively. The difference of the remaining 9 models is all below 3, indicating a high overall accuracy.

3.2. Consistency comparison between model detection results and manual counting results

The detection results of the 12 object detection models are consistent with the results counted by workers ([Fig. 5](#) and [Fig. 6](#)). Except for YOLOv4 and YOLOv7tiny, whose R^2 values in GSF detection are around 0.94, the correlation coefficients between the remaining models and manual detection results in GSF recognition are all above 0.95. In GVF recognition, the correlation coefficients between YOLOv8n, YOLOv4 and YOLOv7, which have the lowest correlation with manual detection results, are 0.9702, 0.971 and 0.9748, respectively, while the results of the remaining 9 models are around 0.98, indicating that the number of GSF and GVF detected by the

Table 2

Comparison of the performance of different models on the validation set.

Category	mAP	AP-GSF	AP-GVF	Model size (MB)	Training Time (h)	GFLOPs	Rank
YOLOv3	94.20%	96.70%	91.70%	123.40	10.57	154.60	8
YOLOv4	92.80%	94.80%	90.90%	105.50	10.77	118.90	11
YOLO-R	94.30%	96.50%	92.00%	105.50	9.95	118.90	6
YOLOv5n	94.70%	96.00%	93.40%	3.80	2.61	4.10	1
YOLOv5m	94.90%	96.70%	93.10%	42.10	3.30	47.90	3
YOLOv5x	95.00%	96.80%	93.20%	173.00	5.97	203.80	7
YOLOv7tiny	90.50%	92.00%	89.10%	12.30	7.93	13.00	12
YOLOv7	94.60%	97.00%	92.20%	74.80	9.55	103.20	5
YOLOv7x	94.40%	96.80%	92.00%	142.10	14.97	188.00	10
YOLOv8n	94.90%	96.40%	93.30%	6.30	1.45	8.10	2
YOLOv8m	94.00%	94.90%	93.00%	52.00	2.94	78.70	4
YOLOv8x	94.10%	95.60%	92.60%	136.70	6.38	257.40	9

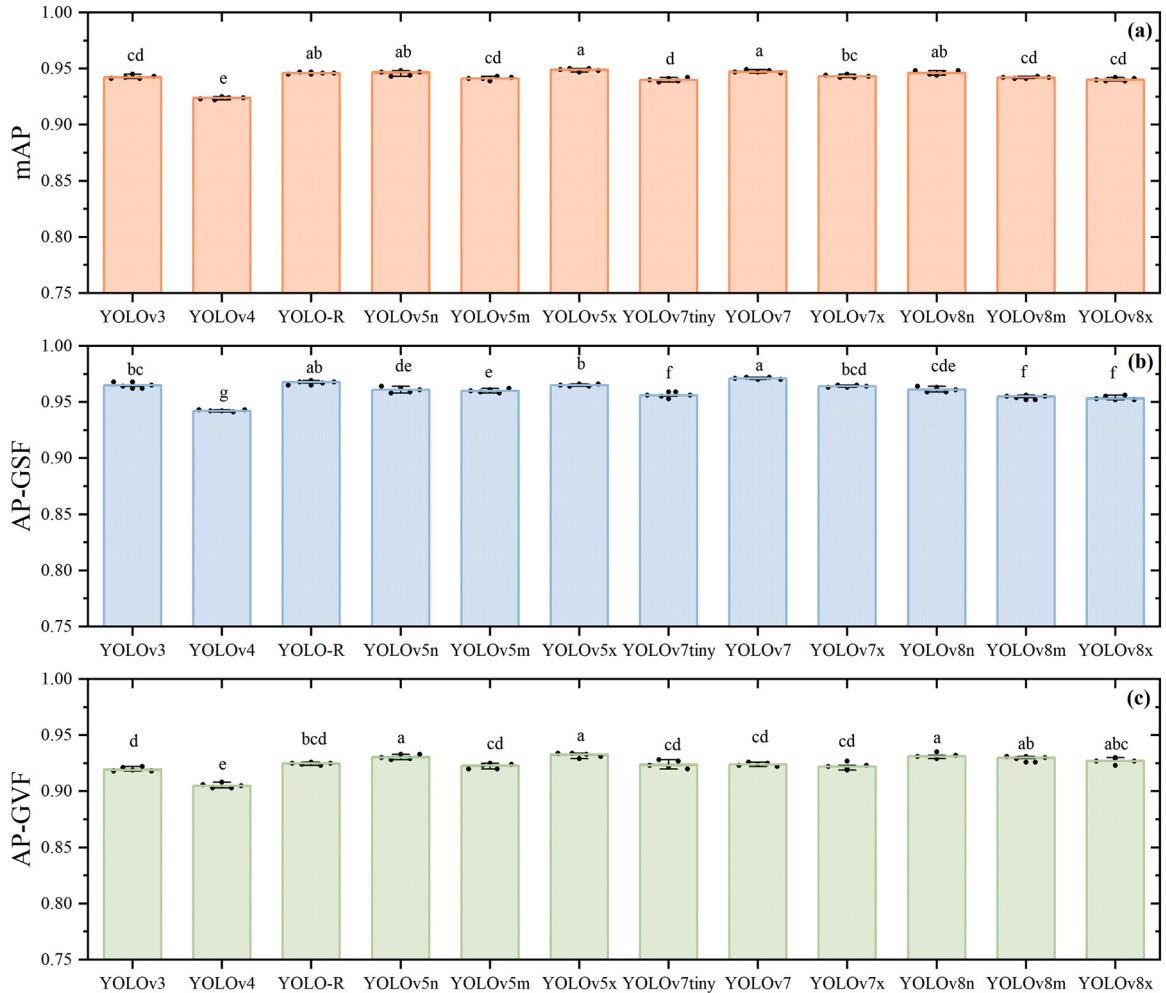


Fig. 3. Using mAP (a), AP-GSF (b) and AP-GVF (c) to take robustness evaluation based on 5-fold cross-validation. Values with different letters significantly differ ($p < 0.05$).

object detection model has practical reference value.

In terms of detection time, when using CPU I9 10900X for detection, YOLOv8n and YOLOv5n completed the detection of this dataset composed of 225 photos in only 12.91 seconds and 14.78 seconds respectively, saving more than half of the time compared to third place YOLOv7tiny which took 33.98 s, and maintained stable accuracy performance. The longest time spent by YOLOv5x and YOLOv7x was 155.97 s and 176.10 s, respectively. In contrast, trained workers need about 156 minutes to count GSF and GVF in this set. In this way, even the most complex YOLOv7x has a detection efficiency that is 52 times higher than manual counting. In a situation where this detection accuracy can be maintained, even using the most complex YOLOv7x is enough to make the combination of UAVRS and deep learning object detection models show great efficiency advantages over traditional ground survey methods in large-scale field detection tasks.

3.3. Quantitative detection and yield estimation of GSF and GVF within the research area

Fig. 7 shows that there is a good linear relationship between the number of both types of plant flowers and the weight of medicinal materials that can be obtained. According to the obtained linear relationship, harvesting 1000 GSF and GVF can respectively yield 151.38 g and 162.14 g of medicinal materials. This result can also explain part of the reason for the depletion of GSF and GVF resources: because the unit weight is too light, people often need to carry out comprehensive harvesting in the same population habitat to reduce the economic expenditure and efficiency loss caused by traveling to different habitats.

In our study, the results of GSF and GVF detection using the same model on 3 devices equipped with different CPUs were consistent (Fig. 8), while the number of two types of flowers detected by different models had some differences. Regarding GSF quantity detection, YOLOv8m and YOLOv5x detected 36,777 and 36,854 GSF, respectively, which were the models with the least detected GSF. In contrast, YOLOv4 and YOLOv7tiny detected 41,123 and 40,277 GSF, respectively. According to the detection results of the object



Fig. 4. Example of detection results from different YOLO models.

detection model, the total yield of dried GSF that can be harvested as medicinal materials in this area ranges from 5.57 kg to 6.22 kg, about 1.88–2.10 g·m⁻². Regarding GVF detection, the two models that detected the least GVF were still YOLOv5m and YOLOv8x, with detected GVF numbers of 80,607 and 80,621, respectively. In comparison, the two models that detected the most GVF were YOLOv4 and YOLO-R, with detected GVF numbers of 84,344 and 83,393, respectively. Therefore, the yield range of dried GVF that can be harvested as medicinal materials in this area is expected to be 13.07 kg to 13.67 kg, about 4.42–4.62 g·m⁻².

Although the detection accuracy of different models is not affected by hardware conditions, their detection efficiency is

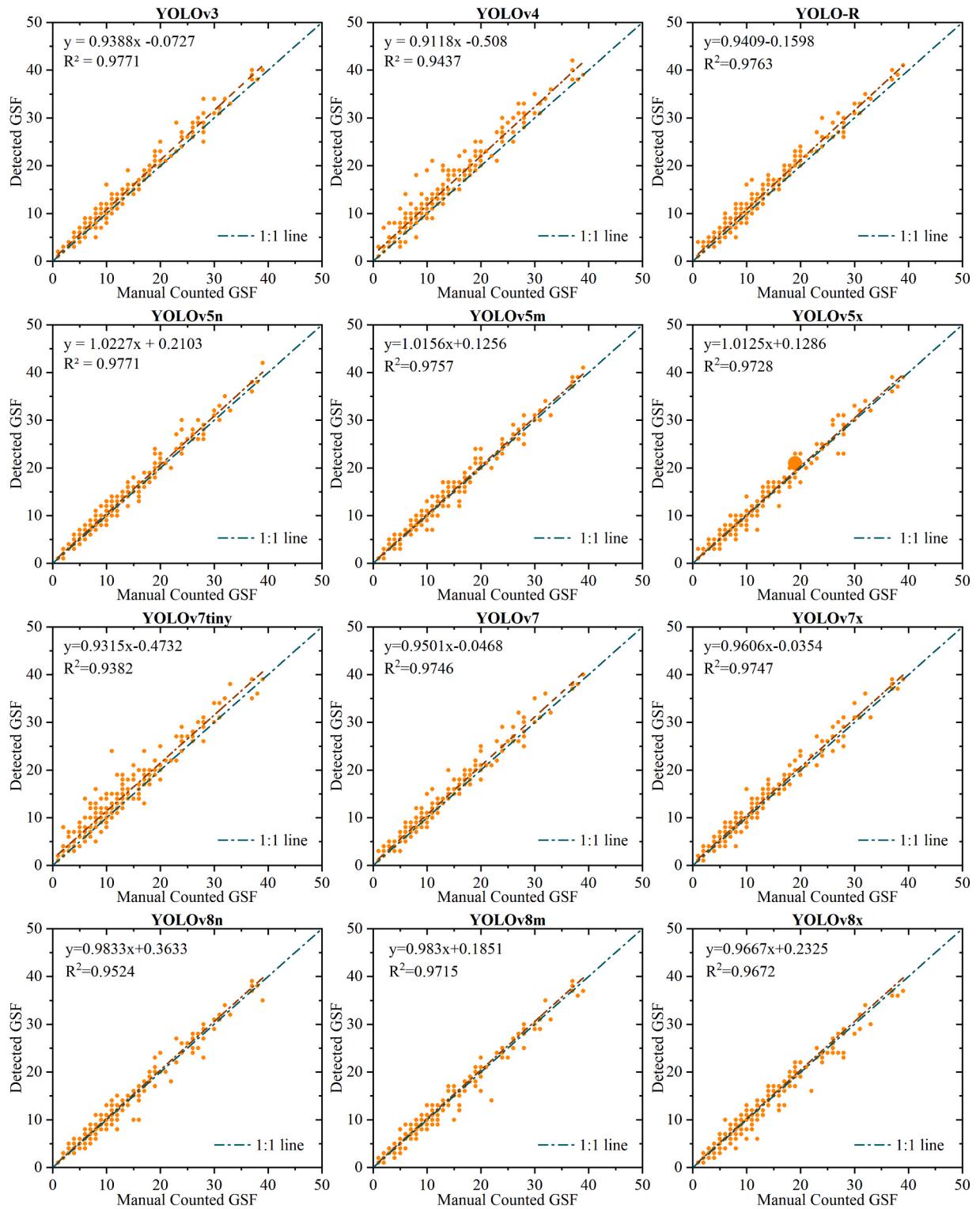


Fig. 5. Consistency comparison between model detection results and manual counting results (GSF).

significantly impacted (Fig. 9). YOLOv5n and YOLOv8n are the fastest two models on all 3 devices, but there is a significant difference in the time it takes, at 265.40 s, 272.3 s, and 409.33 s of YOLOv5n, respectively, and at 323.52 s, 251.77 s and 374.47 s of YOLOv8n, respectively. The difference in time for YOLOv5x, which takes the longest to complete detection of the entire dataset on three devices,

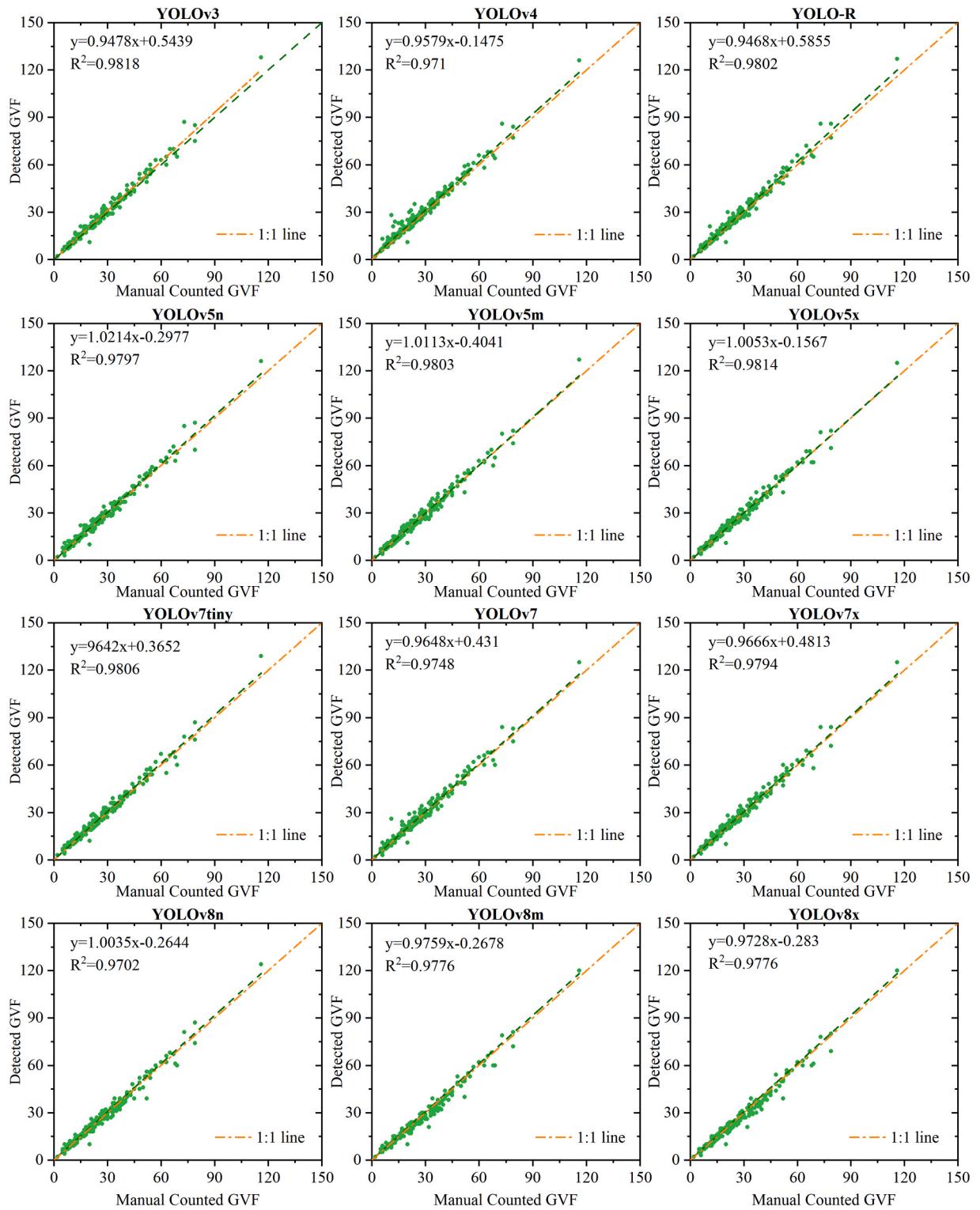


Fig. 6. Consistency comparison between model detection results and manual counting results (GVF).

is even more pronounced at 2464.44 s, 3300.31 s, and 7749.33 s, respectively. This indicates that different hardware levels can significantly determine the time required for detection. From the perspective of Frames per second (FPS), even the fastest YOLOv5n gets an FPS of only 12.31, 11.99, and 7.98 on 3 devices, respectively. Although it has a clear advantage over manual detection

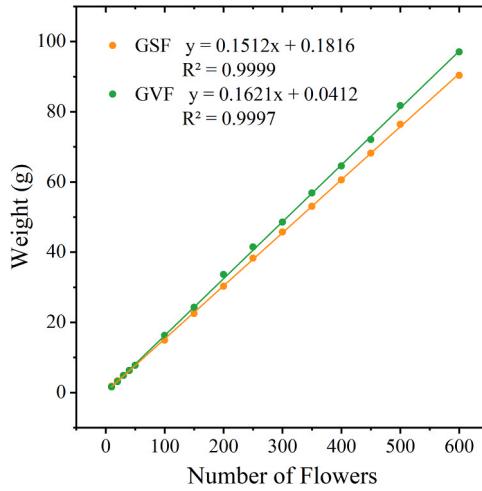


Fig. 7. Relationship between the number and weight of dried GSF and GVF.

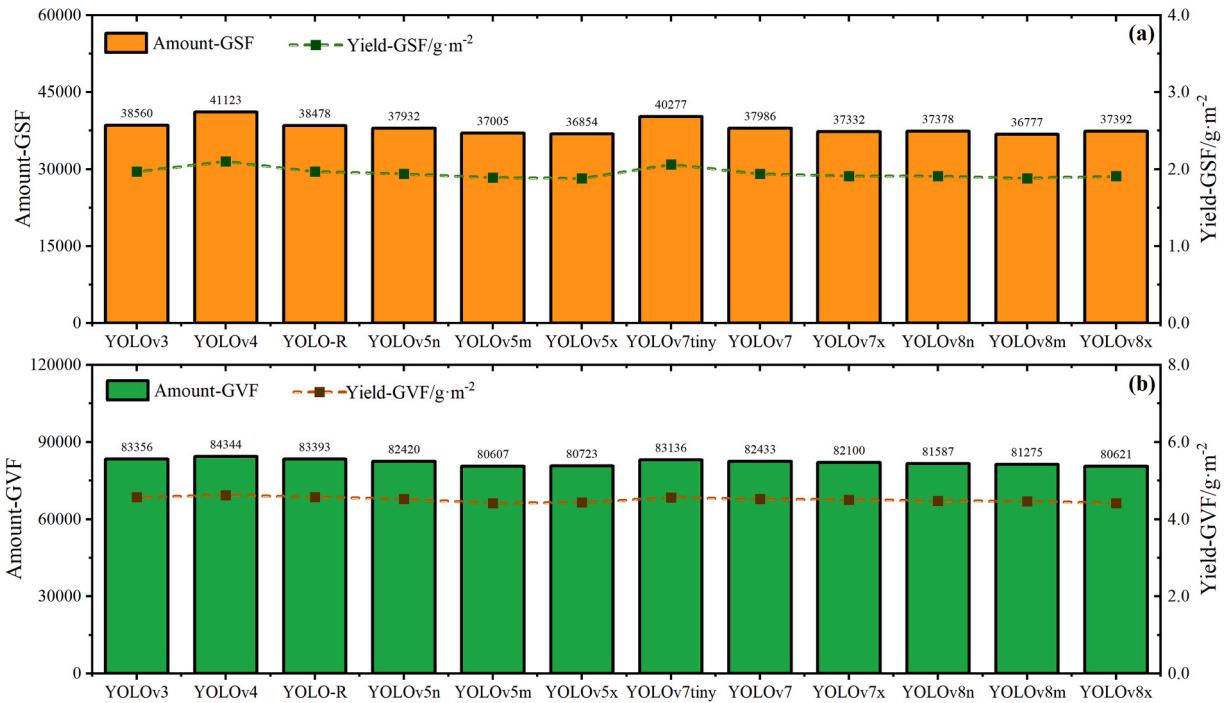


Fig. 8. Results of GSF (a) and GVF (b) quantity detection and yield evaluation in the study area.

efficiency, edge devices that rely solely on the CPU for calculation are still far from the ideal detection goal of real-time detection, so this means that it is necessary to detect 24 photos per second.

Fig. 10 uses box plots and normal curves to show the distribution of GSF and GVF detected by traversing the images of the study area using 12 models. In the results detected by YOLOv5x, the number of GSF ranges from 0 to 52, and the first and third quartiles are 6 and 15, respectively (Fig. 10a). Compared with GSF, GVF has a higher abundance level, with a range of 0–132 and first and third quartiles of 14 and 32 respectively (Fig. 10b). The situation of the other 11 models is similar to that of YOLOv5x, once again confirming that there is no significant difference in accuracy among the 12 models.

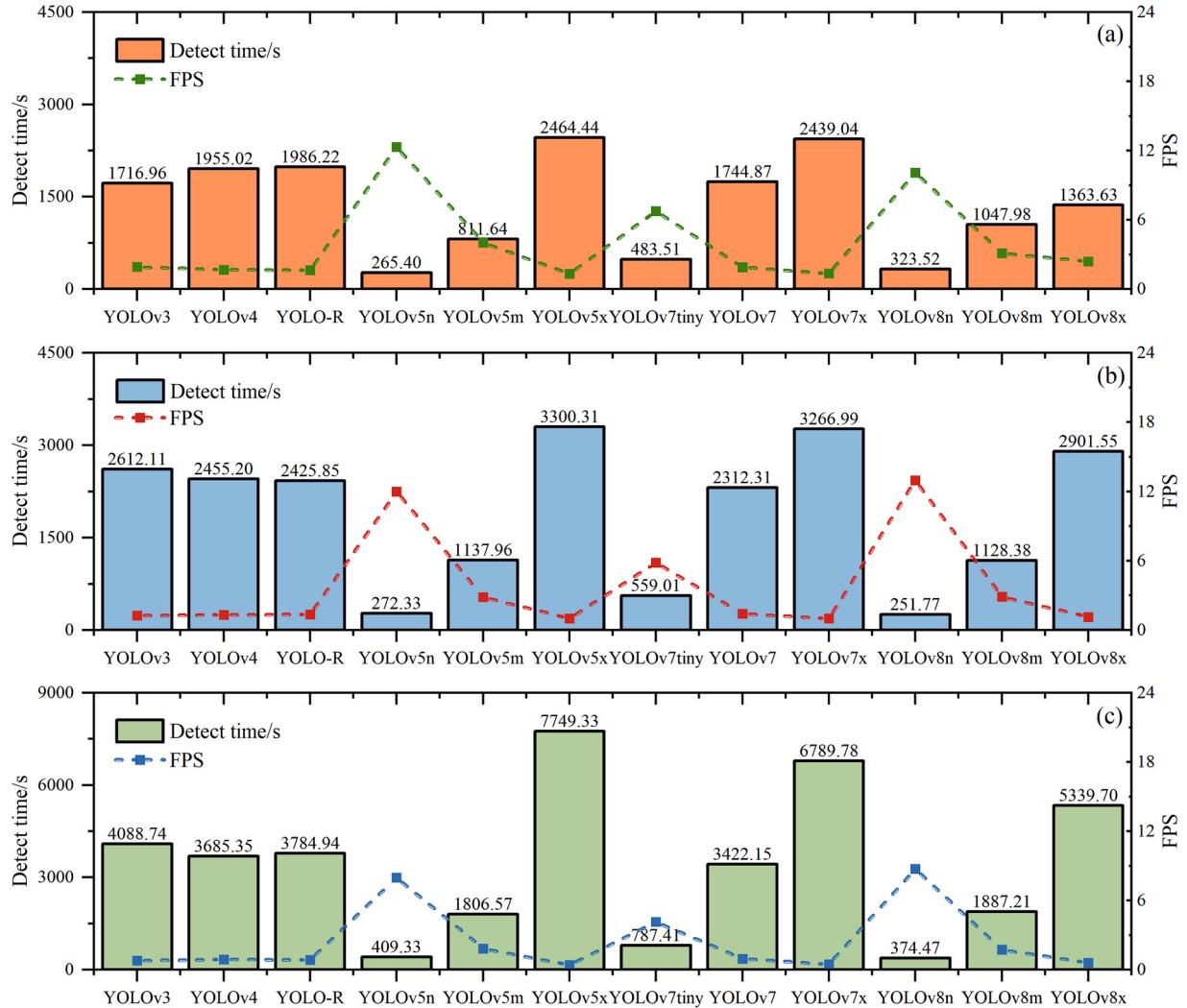


Fig. 9. The 12 YOLO models' performance on 3 devices to traverse the dataset PC1 (a), PC2 (b), and Laptop1 (c).

4. Discussion

4.1. The feasibility of detecting GSF and GVF using UAVRS and deep learning

Many medicinal plants live on the meadows that cover most of the Qinghai-Tibet Plateau, and many of them are facing the threat from both climate change and soaring human activities (Zhao et al., 2022), and their habitats are often far from the limited road coverage. Driving to these habitats can cause damage to the ecology along the way. At the same time, due to the labor-intensive nature of traditional field surveys, conducting surveys in these areas is also a challenge in terms of human resources, and this survey behavior will further increase the pressure on rare plant habitats. UAVRS provides a new research idea for the investigation of rare plants, not only greatly reducing the labor burden of workers, but also reducing the environmental hazards of the survey area to a minimum with non-contact survey methods.

Obtaining efficient and high-precision survey results of wild medicinal plant resources in ecologically fragile areas such as the Qinghai-Tibet Plateau, while minimizing damage to the environment, is a challenging problem. This study aims to verify the feasibility of using UAVRS combined with deep learning object detection algorithms for wild medicinal plant resource surveys, which has practical significance. In addition, by examining the 12 YOLO models from the perspectives of accuracy and efficiency, selecting variants that can maintain satisfactory accuracy and run smoothly on mobile devices with low computing performance can provide a reference for real field survey scenarios.

Firstly, as shown in Table 2, Fig. 3, and Fig. 8, although YOLOv5x slightly outperforms the other 8 object detection models in terms of accuracy indicators, with detection accuracies of 96.80% and 93.20% for GSF and GVF in the test set, respectively. However, the detection accuracy of the remaining 11 models can also reach a high level, with the gap with YOLOv5x all within 1.00%. Considering

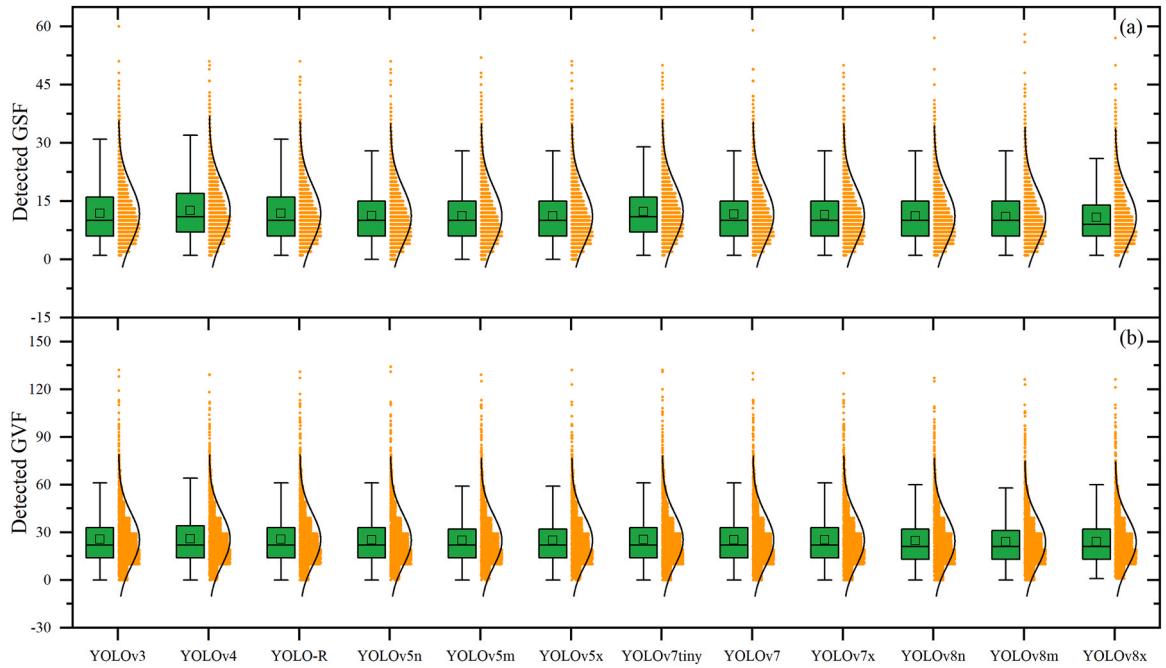


Fig. 10. The distribution of the number of GSF (a) and GVF (b) in every single image.

that YOLOv5n and YOLOv8n achieve a precision level that is almost the same as that of YOLOv5x with a simpler model structure and fewer parameter quantities, YOLOv5n and YOLOv8n are undoubtedly more attractive choices.

4.2. Limitations encountered in this study

Although our research results show that combining deep learning and UAVRS for assessing plateau medicinal plant resources is worth promoting to other endangered plant resource surveys, however, there are still some limiting factors in this resource assessment. Firstly, due to the small size of GSF and GVF and the limited focal length of the camera we used, it is necessary to obtain images at lower flight altitudes, which not only reduces work efficiency but also makes terrain fluctuations more significant on ground sampling rates. Therefore, there may be a large degree of dispersion in the geometric features of flowers in different regions, which may increase the difficulty of fine-tuning object detection models or even cause a higher probability of false detection. This will be more obvious in more complex terrain conditions. Therefore, configuring terrain-following functions for UAVRS or upgrading to cameras with more pixels, and longer focal length to acquire a wider field of view, may eliminate some of the impact of terrain changes while ensuring that the resolution is sufficient for object detection models to learn key features and improve image quality consistency (Wubben et al., 2022).

Besides, *Gentians* often only flower in sunny weather, but due to the plateau area's climate's complexity, field surveys under ideal sunny and windless conditions are often challenging. Strong gusts and dramatic changes in sunlight illumination often accompany these surveys, making obtaining stable quality UAVRS images difficult. This study attempted to reduce motion blur caused by wind on flowers by decreasing the shutter speed during the UAVRS shooting process. Additionally, we adjusted the exposure of photos through data enhancement to simulate images under different weather conditions. However, these limitations still need to be addressed by obtaining an image dataset under more complex weather conditions to verify this study's robustness.

What's more, many closely related gentians occur in the same locality, for example, *G. veitchiorum* co-occur with *G. lawrencei* and *G. futtereri*, *G. szechenyi* co-occur with *G. stipitata*. In addition, *G. veitchiorum* and *G. szechenyi* often occur with genera likely *Aster*, *Anaphalis* and *Leontopodium* which may interfere the detection, but constrain by limited research grant and experience of researchers, these disturb issues have not been thoroughly explored. Thus, to further refine the whole workflow, validation experiments in more localities having closely related gentians are considered important to pave the road of using this material in large-scale resource surveys.

4.3. Application for monitoring *G. szechenyi* flowers and *G. veitchiorum* flowers

Deep learning will change the way plant resource survey are conducted profoundly, because of its objective, non-contact and easy to deploy. Resource survey methods that primarily use sampling as a research method often have a fixed error between the recorded plant abundance and the actual situation due to the recorder's subjective influence. This phenomenon is more commonly known as observer error (Morrison 2016; Morrison and Young 2016). More dangerously, in large-scale survey tasks, due to factors such as reduced concentration of measurement personnel and fluctuations in measurement standards between different time periods or

different measurement personnel, the credibility of survey results is reduced. However, the object detection algorithm used in this study can perform flower detection work with stable detection standards and always maintain satisfactory accuracy and high statistical consistency. Moreover, UAVRS images obtained from the study area can be stored all the time and data can be traced back at any time. Subsequent developed algorithms can also be easily applied to avoid unnecessary data collection and further reduce research costs.

Via monitoring *G. szechuenii* flowers and *G. veitchiorum* flowers, our derived data can have rich application scenarios. For example, since the geographic coordinates can be easily extracted and brought into GIS, population distribution and dynamics can be monitored to ensure sustainable harvesting which presents a potential solution to address the endangered status and propensity for over-harvesting, with which study on population reproductive strategies, optimal harvesting models can get more comprehensive data. Further, it can be combined with harvesting robots to perform automated harvesting based on optimal harvesting strategies to achieve sustainable development of plateau medicinal plant populations. Besides, monitoring the population characteristics of different habitats with varying climates, geographies, terrains, and community features, we may provide new perspectives on the migration and change of populations, as well as offer data support to help increase the understanding of *Gentiana* plants in the Qinghai-Tibetan Plateau. What's more, when using Metashape to build orthorectified mosaics, the generated 3D models can provide high-precision slope and aspect data for species distribution models (SDM) (Elith and Leathwick 2009), which may help to discover more wild populations of *Gentiana* plants, with which we can better evaluate the resource abundance and even the endangered status of the population.

5. Conclusions

Deep learning and unmanned aerial vehicle remote sensing (UAVRS) have the potential to become important tools for surveying wild plants. However, the ease of use for intensive survey tasks in the field needs further investigation. In this study, we compared the You Only Look Once (YOLO) series models, known for their detection speed, to identify *G. szechuenii* flower (GSF) and *G. veitchiorum* flower (GVF) in UAVRS images. We also compared different models' accuracy and efficiency performance on edge computing devices with varying performance levels. Compared with traditional ground survey methods, deep learning methods have higher accuracy, stronger stability, greater survey efficiency, and lower survey costs. The YOLOv5n and YOLOv8n models showed high accuracy and detection speed in GSF and GVF detection, indicating their potential for development in field plant surveys. By obtaining more extensive datasets involving wild populations through UAVRS and further improving the model for the target, this method can provide data support for more reasonable harvesting suggestions. This will help improve the over-harvesting of *G. szechuenii* and *G. veitchiorum*, two threatened wild medicinal plants. With further application of deep learning and UAVRS, this method can be applied to a broader range of wild plant population surveys, promoting the development of wild plant conservation.

CRediT authorship contribution statement

Shihong Zhong: Supervision, Project administration. **Rui Gu:** Writing – review & editing, Resources, Funding acquisition. **Yu Liang:** Validation. **Jiawei Luo:** Software. **Rong Ding:** Validation. **Ziyi Li:** Methodology, Investigation. **Chenghui Wang:** Writing – original draft, Visualization, Formal analysis, Conceptualization.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Rui Gu reports financial support was provided by National Natural Science Foundation of China. Shihong Zhong reports financial support was provided by Science and Technology Department of Sichuan Province. Shihong Zhong reports financial support was provided by Southwest Minzu University

Data availability

Data will be made available on request.

Acknowledgments

The research is sponsored in part by grants from the Natural Science Foundation of China (Grant No. 82073964). The project was also supported by Key R&D Projects of Sichuan Science and Technology Department (Grant No. 20ZDYF2376) and Nationalities Introduces Talented Research Start-up Project of Southwest Minzu University (Grant No. RQD2021055).

References

- Adar, S., Sternberg, M., Paz-Kagan, T., Henkin, Z., Dovrat, G., Zaady, E., Argaman, E., 2022. Estimation of aboveground biomass production using an unmanned aerial vehicle (UAV) and VENeS satellite imagery in Mediterranean and semiarid rangelands. *Remote Sens. Appl.: Soc. Environ.* **26**, 100753.
- Beamish, A., Raynolds, M.K., Epstein, H., Frost, G.V., Macander, M.J., Bergstedt, H., Bartsch, A., Kruse, S., Miles, V., Tanis, C.M., Heim, B., Fuchs, M., Chabrilat, S., Shevtsova, I., Verdonen, M., Wagner, J., 2020. Recent trends and remaining challenges for optical remote sensing of Arctic tundra vegetation: a review and outlook. *Remote Sens. Environ.* **246**, 111872.
- Bochkovskiy, A., Wang, C.-Y., Liao, H.-Y.M., 2020. Yolov4: Optimal speed and accuracy of object detection. *arXiv Prepr. arXiv 2004, 10934*.

- Bomantara, Y.A., Mustafa, H., Bartholomeus, H., Kooistra, L., 2023. Detection of artificial seed-like objects from UAV Imagery. *Remote Sens.* **15**, 1637.
- Chen, G., Kéry, M., Plattner, M., Ma, K., Gardner, B., Zuidema, P., 2013. Imperfect detection is the rule rather than the exception in plant distribution studies. *J. Ecol.* **101**, 183–191.
- Chi, X., Zhang, F., Gao, Q., Xing, R., Chen, S., 2021. A review on the ethnomedicinal usage, phytochemistry, and pharmacological properties of Gentianaeae (Gentianaceae) in Tibetan medicine. *Plants* **10**, 2383.
- Chi, X., Zhang, Z., Xu, X., Zhang, X., Zhao, Z., Liu, Y., Wang, Q., Wang, H., Li, Y., Yang, G., Guo, L., Tang, Z., Huang, L., 2017. Threatened medicinal plants in China: Distributions and conservation priorities. *Biol. Conserv.* **210**, 89–95.
- Coffer, M.M., Graybill, D.D., Whitman, P.J., Schaeffer, B.A., Salls, W.B., Zimmerman, R.C., Hill, V., Lebrasse, M.C., Li, J., Keith, D.J., Kaldy, J., Colarusso, P., Raulerson, G., Ward, D., Kenworthy, W.J., 2023. Providing a framework for seagrass mapping in United States coastal ecosystems using high spatial resolution satellite imagery. *J. Environ. Manag.* **337**, 117669.
- Colomina, I., Molina, P., 2014. Unmanned aerial systems for photogrammetry and remote sensing: a review. *ISPRS J. Photogramm. Remote Sens.* **92**, 79–97.
- Committee, F.O.C.E. (2018). *Flora of China. Flora of China*.
- Corlett, R.T., 2016. Plant diversity in a changing world: status, trends, and conservation needs. *Plant Divers.* **38**, 10–16.
- Corlett, R.T., 2023. Achieving zero extinction for land plants. *Trends Plant Sci.* **28**, 913–923.
- da Silva, S.D.P., Eugenio, F.C., Fantinel, R.A., Amaral, Ld.P., dos Santos, A.R., Mallmann, C.L., dos Santos, F.D., Pereira, R.S., Ruoso, R., 2023. Modeling and detection of invasive trees using UAV image and machine learning in a subtropical forest in Brazil. *Ecol. Inform.* **74**, 101989.
- Dash, J.P., Watt, M.S., Paul, T.S.H., Morgenroth, J., & Pearse, G.D. (2019). Early Detection of Invasive Exotic Trees Using UAV and Manned Aircraft Multispectral and LiDAR Data. In, *Remote Sensing*.
- Ding, R., Luo, J., Wang, C., Yu, L., Yang, J., Wang, M., Zhong, S., & Gu, R. (2023). Identifying and mapping individual medicinal plant *Lamiophlomis rotata* at high elevations by using unmanned aerial vehicles and deep learning. *Plant Methods*.
- Favre, A., Pringle, J.S., Heckenhauer, J., Kozuharova, E., Gao, Q., Lemmon, E.M., Lemmon, A.R., Sun, H., Tkach, N., Gebauer, S., 2020. Phylogenetic relationships and sectional delineation within Gentiana (Gentianaceae). *Taxon* **69**, 1221–1238.
- Fernández-Guisuraga, J.M., Calvo, L., Suárez-Seco, S., 2022. Monitoring post-fire neighborhood competition effects on pine saplings under different environmental conditions by means of UAV multispectral data and structure-from-motion photogrammetry. *J. Environ. Manag.* **305**, 114373.
- Fu, B., He, X., Yao, H., Liang, Y., Deng, T., He, H., Fan, D., Lan, G., He, W., 2022. Comparison of RFE-DL and stacking ensemble learning algorithms for classifying mangrove species on UAV multispectral images. *Int. J. Appl. Earth Obs. Geoinf.* **112**, 102890.
- Fu, H., Sun, G., Zhang, L., Zhang, A., Ren, J., Jia, X., Li, F., 2023. Three-dimensional singular spectrum analysis for precise land cover classification from UAV-borne hyperspectral benchmark datasets. *ISPRS J. Photogramm. Remote Sens.* **203**, 115–134.
- Girshick, R. (2015). Fast r-cnn. In, *Proceedings of the IEEE international conference on computer vision* (pp. 1440–1448).
- Goetttsch, B., Hilton-Taylor, C., Cruz-Piñón, G., Duffy, J.P., Frances, A., Hernández, H.M., Inger, R., Pollock, C., Schipper, J., Superina, M., Taylor, N.P., Tognelli, M., Abba, A.M., Arias, S., Arreola-Nava, H.J., Baker, M.A., Barcenas, R.T., Barrios, D., Braun, P., Butterworth, C.A., Bürquez, A., Cáceres, F., Chazaro-Basañez, M., Corral-Díaz, R., del Valle Perea, M., Demaio, P.H., Duarte de Barros, W.A., Durán, R., Yancas, L.F., Felger, R.S., Fitz-Maurice, B., Fitz-Maurice, W.A., Gann, G., Gómez-Hinostrosa, C., Gonzales-Torres, L.R., Patrick Griffith, M., Guerrero, P.C., Hammel, B., Heil, K.D., Hernández-Oria, J.G., Hoffmann, M., Ishihara, M.I., Kiesling, R., Larocca, J., León-de la Luz, J.L., Loaiza, S., Lowry, C.R., Machado, M., Majure, M.C., Ávalos, L.C., Martorell, J.G.M., Maschinski, C., Méndez, J., Mittermeier, E., Nassar, R.A., Negrón-Ortiz, J.M., Oakley, V., Ortega-Baes, L.J., Ferreira, P., Pinkava, A.B.P., Porter, D.J., Puente-Martinez, J.M., Gamarra, R., Pérez, J.R., Martínez, P.S., Smith, E.S., Manuel Sotomayor, M., del, M., C, J., Stuart, S.N., Muñoz, J.L.T., Terrazas, T., Terry, M., Trevisson, M., Valverde, T., Van Devender, T.R., Véliz-Pérez, M.E., Walter, H.E., Wyatt, S.A., Zappi, D., Alejandro Zavala-Hurtado, J., Gaston, K.J., 2015. High proportion of cactus species threatened with extinction. *Nat. Plants* **1**, 15142.
- Hamilton, S.M., Morris, R.H., Carvalho, R.C., Roder, N., Barlow, P., Mills, K., Wang, L., 2020. Evaluating techniques for mapping island vegetation from unmanned aerial vehicle (UAV) images: pixel classification, visual interpretation and machine learning approaches. *Int. J. Appl. Earth Obs. Geoinf.* **89**, 102085.
- He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In, *Proceedings of the IEEE international conference on computer vision* (pp. 2961–2969).
- Ho, T.-n., & Liu, S.-w (2001). A worldwide monograph of Gentiana. In.
- Hunault-Fontbonne, J., Eyyindson, K., 2023. Bridging the gap between forest planning and ecology in biodiversity forecasts: a review. *Ecol. Indic.* **154**, 110620.
- Ishida, T., Kurihara, J., Viray, F.A., Namuco, S.B., Paringit, E.C., Perez, G.J., Takahashi, Y., Marciano, J.J., 2018. A novel approach for vegetation classification using UAV-based hyperspectral imaging. *Comput. Electron. Agric.* **144**, 80–85.
- Jiménez López, J., & Mulero-Pázmány, M. (2019). Drones for Conservation in Protected Areas: Present and Future. In, *Drones*.
- Jocher, G., Chaurasia, A., & Qiu, J. (2023). YOLO by Ultralytics. URL: (<https://github.com/ultralytics/ultralytics>).
- Jocher, G., Stoken, A., Borovec, J., Changyu, L., Hogan, A., Chaurasia, A., Diaconu, L., Ingham, F., Colmagro, A., & Ye, H. (2021). YOLOv5 by Ultralytics. *Zenodo*.
- Kamilaris, A., Prenafta-Boldú, F.X., 2018. Deep learning in agriculture: a survey. *Comput. Electron. Agric.* **147**, 70–90.
- Kupková, L., Červená, L., Potůčková, M., Lysák, J., Roubalová, M., Hrázský, Z., Březina, S., Epstein, H.E., Müllerová, J., 2023. Towards reliable monitoring of grass species in nature conservation: evaluation of the potential of UAV and PlanetScope multi-temporal data in the Central European tundra. *Remote Sens. Environ.* **294**, 113645.
- Li, C., Li, L., Jiang, H., Weng, K., Geng, Y., Li, L., Ke, Z., Li, Q., Cheng, M., Nie, W., 2022. YOLOv6: a single-stage object detection framework for industrial applications. *arXiv Prepr. arXiv 2209, 02976*.
- Li, Y., Zhang, J., Fan, J.-y., Zhong, S.-h., Gu, R., 2023. Tibetan medicine Bang Jian: a comprehensive review on botanical characterization, traditional use, phytochemistry, and pharmacology. *Front. Pharmacol.* **14**.
- Mafanya, M., Tsele, P., Botai, J., Manyama, P., Swart, B., Monate, T., 2017. Evaluating pixel and object based image classification techniques for mapping plant invasions from UAV derived aerial imagery: Harrisia pomanensis as a case study. *ISPRS J. Photogramm. Remote Sens.* **129**, 1–11.
- Mao, P., Jiang, B., Shi, Z., He, Y., Shen, T., Qiu, G.Y., 2023. Effects of UAV flight height on biomass estimation of desert shrub communities. *Ecol. Indic.* **154**, 110698.
- Matas-Granados, L., Pizarro, M., Cayuela, L., Domingo, D., Gómez, D., García, M.B., 2022. Long-term monitoring of NDVI changes by remote sensing to assess the vulnerability of threatened plants. *Biol. Conserv.* **265**, 109428.
- Nakamura, A., Kitching, R.L., Cao, M., Creedy, T.J., Fayle, T.M., Freiberg, M., Hewitt, C.N., Itioka, T., Koh, L.P., Ma, K., Malhi, Y., Mitchell, A., Novotny, V., Ozanne, C. M.P., Song, L., Wang, H., Ashton, L.A., 2017. Forests and Their Canopies: achievements and Horizons in Canopy Science. *Trends Ecol. Evol.* **32**, 438–451.
- Nan, Y., Zhang, H., Zeng, Y., Zheng, J., Ge, Y., 2023. Intelligent detection of Multi-Class pitaya fruits in target picking row based on WGB-YOLO network. *Comput. Electron. Agric.* **208**, 107780.
- Nunes, Y.R.F., Souza, C.S., Azevedo, I.F.Pd, Oliveira, O.Sd, Frazão, L.A., Fonseca, R.S., Santos, R.Md, Neves, W.V., 2022. Vegetation structure and edaphic factors in veredas reflect different conservation status in these threatened areas. *For. Ecosyst.* **9**, 100036.
- Oh, S., Chang, A., Ashapure, A., Jung, J., Dube, N., Maeda, M., Gonzalez, D., Landivar, J., 2020. Plant Counting of Cotton from UAS Imagery Using Deep Learning-Based Object Detection Framework. *Remote Sens.* **12**, 2981.
- Peng, J., Wang, D., Liao, X., Shao, Q., Sun, Z., Yue, H., Ye, H., 2020. Wild animal survey using UAS imagery and deep learning: modified Faster R-CNN for kiang detection in Tibetan Plateau. *ISPRS J. Photogramm. Remote Sens.* **169**, 364–376.
- Pohjanmies, T., Jašková, A., Hotanen, J.-P., Manninen, O., Salemaa, M., Tolvanen, A., Merilä, P., 2021. Abundance and diversity of edible wild plants in managed boreal forests. *For. Ecol. Manag.* **491**, 119151.
- Puri, V., Nayyar, A., Raja, L., 2017. Agriculture drones: a modern breakthrough in precision agriculture. *J. Stat. Manag. Syst.* **20**, 507–518.
- Redmon, J., Farhadi, A., 2018. Yolov3: An incremental improvement. *arXiv Prepr. arXiv 1804, 02767*.
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In, *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779–788).
- Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster r-cnn: towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* **28**.

- Royimani, L., Mutanga, O., Odindi, J., Dube, T., Matongera, T.N., 2019. Advancements in satellite remote sensing for mapping and monitoring of alien invasive plant species (AIPs). *Phys. Chem. Earth Parts A/B/C.* **112**, 237–245.
- Scheeres, J., de Jong, J., Brede, B., Brancalion, P.H.S., Broadbent, E.N., Zambrano, A.M.A., Gorgens, E.B., Silva, C.A., Valbuena, R., Molin, P., Stark, S., Rodrigues, R.R., Santoro, G.B., Resende, A.F., de Almeida, C.T., de Almeida, D.R.A., 2023. Distinguishing forest types in restored tropical landscapes with UAV-borne LiDAR. *Remote Sens. Environ.* **290**, 113533.
- Sotille, M.E., Bremer, U.F., Vieira, G., Velho, L.F., Petsch, C., Auger, J.D., Simões, J.C., 2022. UAV-based classification of maritime Antarctic vegetation types using GEOBIA and random forest. *Ecol. Inform.* **71**, 101768.
- Su, J., Yi, D., Coombes, M., Liu, C., Zhai, X., McDonald-Maier, K., Chen, W.-H., 2022. Spectral analysis and mapping of blackgrass weed by leveraging machine learning and UAV multispectral imagery. *Comput. Electron. Agric.* **192**, 106621.
- Theron, K.J., Pryke, J.S., Latte, N., Samways, M.J., 2022. Mapping an alien invasive shrub within conservation corridors using super-resolution satellite imagery. *J. Environ. Manag.* **321**, 116023.
- Trujillo-Jiménez, M.A., Liberoff, A.L., Pessacq, N., Pacheco, C., Díaz, L., Flaherty, S., 2022. SatRed: New classification land use/land cover model based on multi-spectral satellite images and neural networks applied to a semiarid valley of Patagonia. *Remote Sens. Appl.: Soc. Environ.* **26**, 100703.
- Valente, J., Hiremath, S., Ariza-Sentís, M., Doldersum, M., Kooistra, L., 2022. Mapping of Rumex obtusifolius in nature conservation areas using very high resolution UAV imagery and deep learning. *Int. J. Appl. Earth Obs. Geoinf.* **112**, 102864.
- van Andel, T., van der Hoorn, B., Stech, M., Arostegui, S.B., Miller, J., 2016. A quantitative assessment of the vegetation types on the island of St. Eustatius, Dutch Caribbean. *Glob. Ecol. Conserv.* **7**, 59–69.
- VonBank, J.A., Casper, A.F., Yetter, A.P., Hagy, H.M., 2017. Evaluating a rapid aerial survey for floating-leaved aquatic vegetation. *Wetlands* **37**, 753–762.
- Wan Nurazwin Syazwani, R., Muhammad Asraf, H., Megat Syahirul Amin, M.A., Nur Dalila, K.A., 2022. Automated image identification, detection and fruit counting of top-view pineapple crown using machine learning. *Alex. Eng. J.* **61**, 1265–1276.
- Wang, C.-Y., Yeh, I.-H., Liao, H.-Y.M., 2021. You only learn one representation: Unified network for multiple tasks. *arXiv Prepr. arXiv 2105, 04206*.
- Wang, C.-Y., Bochkovskiy, A., & Liao, H.-Y.M. (2023). 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* (pp. 7464–7475).
- Wolff, F., Kolari, H.M., Villaslada, T., Tahvanainen, M., Korpelainen, T., A. P. P., Zamboni, P., Kumpula, T., 2023. RGB vs. Multispectral imagery: mapping aapa mire plant communities with UAVs. *Ecol. Indic.* **148**, 110140.
- Zhang, C., Atkinson, P.M., George, C., Wen, Z., Diazgranados, M., Gerard, F., 2020. Identifying and mapping individual plants in a highly diverse high-elevation ecosystem using UAV imagery and deep learning. *ISPRS J. Photogramm. Remote Sens.* **169**, 280–291.
- Zhang, T., Vail, S., Duddu, H.S.N., Parkin, I.A.P., Guo, X., Johnson, E.N., Shirtliffe, S.J., 2021. Phenotyping flowering in canola (*Brassica napus* L.) and estimating seed yield using an unmanned aerial vehicle-based imagery. *Front. Plant Sci.* **12**.
- Zhang, Z., Dong, X., Tian, J., Tian, Q., Xi, Y., He, D., 2022. Stand density estimation based on fractional vegetation coverage from Sentinel-2 satellite imagery. *Int. J. Appl. Earth Obs. Geoinf.* **108**, 102760.
- Zhang, Z., Lu, L., Zhao, Y., Wang, Y., Wei, D., Wu, X., Ma, X., 2023. Recent advances in using Chinese Earth observation satellites for remote sensing of vegetation. *ISPRS J. Photogramm. Remote Sens.* **195**, 393–407.
- Zhang, Z.-F., Liu, Y., Lu, L.-Y., Luo, P., 2014. Hepatoprotective activity of Gentiana veitchiorum Hemsl. against carbon tetrachloride-induced hepatotoxicity in mice. *Chin. J. Nat. Med.* **12**, 488–494.
- Zhao, F., Wu, X., Wang, S., 2020. Object-oriented vegetation classification method based on UAV and satellite image fusion. *Procedia Comput. Sci.* **174**, 609–615.
- Zhao, R., Xu, S., Song, P., Zhou, X., Zhang, Y., Yuan, Y., 2022. Distribution patterns of medicinal plant diversity and their conservation priorities in the Qinghai-Tibet Plateau. *Biodiversity Sci.* **30**, 21385.