Automation and Machine Learning (2026)
Clausius Scientific Press, Canada

DOI: 10.23977/autml.2026.070116
ISSN 2516-5003 Vol. 7 Num. 1

A Study on Automatic Multi-Object Detection in Forest Areas Based on UAV Imagery

Yi Ren[1,a], Yongxin Yan[1,b], Yanhua Zhang[1,c], Suli Li[1,d], Dong Jing[1,e], Tianliang Zhang[2,3,f,*]

[1]Weichang Manchu and Mongol Autonomous County State-Owned Luanhe Forest Farm, Weichang County, Chengde, China
[2]Institute of Applied Mathematics, Hebei Academy of Sciences, No. 46 South Youyi Street, Shijiazhuang, China
[3]Hebei Information Security Certification Technology Innovation Center, No. 46 South Youyi Street, Shijiazhuang, China
[a]skq202505@163.com, [b]18830419090@qq.com, [c]928634665@qq.com, [d]867965226@qq.com, [e]13932484169@163.com, [f]tianliangzn@126.com
**Corresponding author*

Keywords: UAV remote sensing; YOLOv5; multi-object detection in forest areas; smart forestry

Abstract: This study focuses on the Saihanba Mechanical Forest Farm in Hebei Province, China, and develops an automated multi-object detection system for forest areas—including graves, cow, and vehicles—based on visible-light imagery from consumer-grade drones and the YOLOv5 deep learning model. The system addresses core operational needs such as fire control during grave-side rituals, ecological monitoring of free-range cow herds, and safety patrols of vehicles in forest areas. A DJI Mavic 3 drone was used to acquire visible-light imagery of the forest area at various flight altitudes. A specialized multi-object dataset for forest areas was constructed, comprising 470 valid images and 1,467 annotated samples, covering the three target categories—graves, cow, and vehicles—across different scenarios. Data augmentation was performed through geometric transformations and mixed augmentation to enhance the model's generalization ability; To address the issues of low detection accuracy for small targets and feature loss in complex forest backgrounds, the YOLOv5 model was enhanced in multiple dimensions. This included adopting the GhostConv lightweight backbone network, embedding a Coordinate Attention (CA) mechanism, replacing the BiFPN with a weighted bidirectional feature fusion structure, and using the EIoU loss function to optimize bounding box regression. Model training was completed via transfer learning using COCO pre-training weights.Experimental results show that the improved YOLOv5 model achieves an average accuracy of 97.25% on the test set, capable of accurately identifying three types of targets in complex forest backgrounds, with excellent resistance to occlusion and background interference. This study validates the feasibility and superiority of combining consumer-grade drone visible-light imagery with deep learning models for automated multi-target monitoring in forest areas. It can provide efficient, low-cost technical support for forest fire early warning, forestry ecological management, and forest safety supervision, and holds significant application potential.

1. Introduction

With the ongoing advancement of smart forestry initiatives, unmanned aerial vehicle (UAV) remote sensing technology has become a core technical tool for forest resource monitoring, pest and disease control, and fire hazard detection, thanks to its advantages of mobility, high spatial and temporal resolution, minimal terrain constraints, and low operational costs[1]. China's forested areas are predominantly located in mountainous and hilly regions characterized by complex topography and dense vegetation cover. Three key monitoring targets are of critical importance to forestry management: scattered gravesites in forested areas pose a major risk of forest fires; illegal use of fire during ancestral worship periods such as Qingming Festival and the Spring Festival is the primary cause of major and catastrophic forest fires; and traditional manual inspections struggle to achieve precise, real-time monitoring of a large number of gravesite locations; Second, overgrazing by free-roaming cow in forest areas damages young trees and saplings as well as surface vegetation, exacerbates soil erosion, and undermines the effectiveness of forest ecological restoration. Manual supervision struggles to track the spatial distribution and movement patterns of cow in real time; Third, the dynamic monitoring of unauthorized vehicles entering forest areas (for illegal logging or unauthorized use of fire) and forest fire emergency vehicles is a critical component of forest safety management. Traditional ground-based monitoring points have limited coverage and cannot achieve comprehensive, real-time patrols across the entire area.

In recent years, the integration of deep learning-based object detection algorithms with UAV remote sensing has yielded significant research results in fields such as agricultural pest and disease identification, wildlife monitoring, and traffic management[2–5]. In the field of livestock and poultry monitoring, Zhang Liwen et al.[6] proposed the PigsTrack tracker to achieve multi-object tracking of group-reared pigs, while Zhao Yongxiang et al.[7] implemented drone tracking and localization of dairy cows based on an improved CenterTrack algorithm, demonstrating the superior recognition capabilities of deep learning models for livestock and poultry targets in complex backgrounds; In vehicle detection, Zang Ke[8] and Xiang Yutao et al.[9] independently performed lightweight modifications on the YOLOv5 model, achieving real-time vehicle detection in UAV remote sensing imagery, thereby providing a technical reference for vehicle target recognition in forested areas.

Existing research has primarily focused on single-target detection in forest areas or single operational scenarios, with limited studies on the joint multi-target detection of three heterogeneous targets—graves, cow, and vehicles—in complex forest backgrounds. Additionally, forest scenes are characterized by severe vegetation occlusion, significant variations in target scale, and high similarity between background and target textures. The original YOLOv5 model suffers from issues such as the loss of small-target features and high false positive rates in complex backgrounds. To address these needs and technical bottlenecks, this study performs multi-dimensional lightweighting and accuracy optimization on the YOLOv5 model. It constructs a multi-target automatic detection method for forest areas based on UAV visible-light imagery, systematically completing data collection, dataset construction, model improvement, and performance validation, thereby providing a technical solution for integrated multi-target monitoring in smart forestry.

2. Materials and Methods

2.1. Overview of the Study Area

The study area is located in the vicinity of the Saihaba Mechanical Forest Farm in Weichang Manchu-Mongol Autonomous County, Chengde City, Hebei Province, China (116°51′–117°39′E, 42°02′–42°36′N). It lies in the transition zone between the Inner Mongolia Plateau and the Northern Hebei Mountains, at an elevation ranging from 1,010 to 1,940 m. The study area is dominated by

vegetation types such as larch, Mongolian pine, and white birch, with large areas of grassland, shrubland, and ravines interspersed throughout the forest. It serves as a core area for the ecological conservation of planted forests and forest fire prevention. Traditional burial sites are scattered throughout the forest, and surrounding villages and towns engage in livestock grazing activities within the forest area. With forest fire roads and production roads crisscrossing the region, this area provides a typical experimental setting for the detection and research of three types of targets: graves, cow, and vehicles.

2.2. Data Collection and Preprocessing

2.2.1. Collection Equipment

This study utilized the DJI Mavic 3 consumer-grade drone as the data acquisition platform. Equipped with a 1-inch CMOS visible-light imaging sensor featuring 20 million effective pixels and a 24 mm equivalent focal length, it supports lossless image storage in JPG format. With its lightweight design, long battery life, and consistent image quality, it is well-suited for large-scale field operations in forestry. On the ground, the DJI RC Pro remote controller was used to control flight paths and enable real-time image transmission.

2.2.2. Collection Plan

Data collection took place from March to October 2025, covering the critical periods for forest fire prevention in spring and fall, as well as the peak vegetation growth season in summer. The data comprehensively captures target scenes under varying vegetation cover and lighting conditions. To ensure uniform lighting in the imagery and avoid interference from backlighting and large areas of shadow, data collection was conducted on sunny days between 9:00 a.m. and 3:00 p.m., when the sun's altitude angle exceeded 45°. The flight altitude was set at 100 m, corresponding to a ground resolution of 3.2 cm/pixel, with a forward overlap of 80% and a side overlap of 70%, and a flight speed of 5 m/s. A total of three aerial survey flights were completed, covering three typical scenarios in the study area: the concentrated graveyard area, the grazing area, and the main forest roads. A total of 470 raw aerial images were acquired, the actual image capture results are shown in Figure 1.

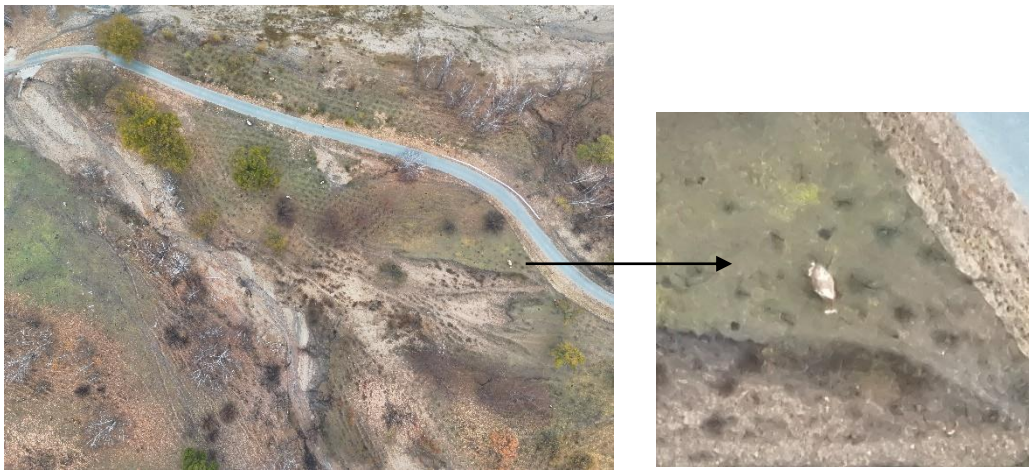

Figure 1: Single Image and Ground Resolution Display

2.2.3. Data Preprocessing

We employ multi-strategy data augmentation methods to expand the dataset and enhance the

model's generalization and robustness. Specific augmentation techniques include: geometric transformation augmentation, which involves randomly rotating images by 90°, 180°, or 270°, performing horizontal or vertical mirroring, and applying random cropping and translation to simulate changes in target appearance under different drone shooting angles and flight altitudes; Pixel transformation augmentation: Adding Gaussian noise to images, randomly adjusting brightness and contrast (±30%), and applying gamma transformations to simulate changes in image features under varying lighting and weather conditions, thereby improving the model's adaptability to complex lighting environments.

Using the above augmentation methods, the dataset was expanded to three times its original size. The preprocessed images were then annotated using the LabelImg tool with bounding boxes, defining three detection classes: `grave` (grave mounds), `cow` (cow), and `car` (vehicles). The final specialized multi-object detection dataset for forest areas was constructed, comprising a total of 470 valid images and 1,467 annotated objects, including 200 grave mounds, 800 cow, and 467 vehicles. The dataset covers target samples of varying sizes, occlusion levels, and background environments, with a minimum target size of 12×15 pixels, and the maximum target size is 380×420 pixels, fully ensuring the diversity and representativeness of the dataset. The dataset was randomly split into training, validation, and test sets in a 7:1.5:1.5 ratio, comprising 329 images in the training set, 71 images in the validation set, and 70 images in the test set. There is no spatial overlap between the datasets, ensuring the objectivity of model performance evaluation.

2.3. Construction of the YOLOv5 Detection Model

2.3.1. Architecture of the Original YOLOv5 Model

YOLOv5 is a classic single-stage object detection model characterized by an excellent balance between detection accuracy and inference speed, as well as high deployment flexibility. In this study, YOLOv5s was selected as the base model. Its overall architecture consists of four components: the input stage, which performs adaptive image resizing, Mosaic data augmentation, and adaptive anchor box calculation; Backbone: Utilizes the CSPDarknet53 architecture, performing image feature extraction via the Focus module, CBS convolutional module, C3 residual module, and SPPF spatial pyramid pooling module; the Neck, which employs an FPN+PAN bidirectional feature fusion architecture to integrate shallow-level texture features with deep-level semantic features; and the Detection Head, which uses a coupled detection head to perform object classification and bounding box regression, outputting the final detection results..

2.3.2. Model Improvement Strategies

To address the challenges of multi-object detection in complex forest environments, this study implements multi-dimensional improvements to the YOLOv5s model. These enhancements aim to improve detection accuracy while reducing the model's computational complexity, thereby meeting the deployment requirements for UAV edge devices. The core improvements are as follows:

Backbone Network Lightweighting: The original convolutions in the C3 module are replaced with GhostConv modules. By generating redundant feature maps through low-cost linear transformations, this approach significantly reduces the number of model parameters and computational load without compromising feature extraction capabilities, thereby improving inference speed. Embedding of Coordinate Attention Mechanism: A coordinate attention (CA) mechanism is embedded into the C3 module of the backbone network. This decomposes channel attention into two one-dimensional feature encoding processes, simultaneously capturing inter-channel dependencies and target location information. This enhances the model's ability to extract features from targets such as cow heads and

small-sized herds while suppressing interference from complex backgrounds like forest vegetation and bare ground. Optimization of the Neck Feature Fusion Network: We replace the original FPN+PAN architecture with a weighted bidirectional feature pyramid network (BiFPN). Through bidirectional cross-scale connections and learnable weights, we perform weighted fusion of feature maps at different resolutions. This addresses the issues of large scale variations among forest targets and the loss of small-target features during transmission, thereby improving the detection accuracy of multi-scale targets. Loss Function Optimization: The original C-IoU loss function is replaced with an E-IoU loss function. While retaining overlap and center-distance losses, this approach directly minimizes the width-to-height ratio difference between the target bounding box and the anchor box. This resolves the slow convergence issue associated with the width-to-height ratio in C-IoU, thereby improving the regression accuracy and convergence speed of the bounding boxes.

2.3.3. Model Training Settings

This study employed a transfer learning approach for model training. The network was initialized using pre-trained weights from the COCO public dataset and then globally fine-tuned using the multi-objective dataset of forest areas collected in this study. This effectively addressed the issues of slow convergence and poor generalization performance typically encountered with small datasets. The model training parameters are set as follows: The input image size is 640×640 pixels, the batch size is 16, and the total number of training epochs is 150. The optimizer used is Stochastic Gradient Descent with Momentum (SGDM), with a momentum value of 0.937, an initial learning rate of 0.01, and a weight decay coefficient of 0.0005. The learning rate is dynamically adjusted using a cosine annealing strategy.

2.4. Model Performance Metrics

This study adopts core evaluation metrics commonly used in the object detection field, including precision (P), recall (R), average precision (AP), and mean average precision (mAP). The formulas for each metric are as follows:

Precision reflects the model's ability to avoid false positives:

$$P = \frac{TP}{TP+FP} \times 100\% \tag{1}$$

Recall reflects the model's ability to avoid false negatives:

$$R = \frac{TP}{TP+FN} \times 100\% \tag{2}$$

Average Precision is the area under the PR curve and comprehensively reflects the detection performance for a single class:

$$AP = \int_0^1 P(R)dR \tag{3}$$

Mean Average Precision is the average of the APs across all classes and serves as a core metric for the model's overall performance:

$$mAP = \frac{1}{n}\sum_{i=1}^{n} AP_i \tag{4}$$

Where: TP represents true positives, i.e., the number of target samples correctly detected by the model; FP represents false positives, i.e., the number of non-target samples incorrectly detected by the model; FN represents false negatives, i.e., the number of true target samples not detected by the model; n represents the total number of detection categories; in this study, n=3.

3. Results and Analysis

3.1. Comparison of Model Detection Performance

We compared the performance of the improved YOLOv5 model with that of the original YOLOv5s model using the test set data; the results are shown in Table 1. As shown in Table 1, the improved YOLOv5 model outperforms the original model in the detection performance of all three object classes, with a comprehensive mAP of 97.25%, representing an increase of 4.11 percentage points over the original model. Among these, the AP for the "Grave" category showed the most significant improvement, increasing by 6.43 percentage points, demonstrating that the improved model possesses stronger feature extraction capabilities for small-sized, low-texture objects in complex backgrounds. The AP for both the 'Cow' and "Vehicle" categories increased by more than 2 percentage points, indicating a significant enhancement in the model's ability to handle occlusion and background interference.

Table 1: Comparison of Performance Between the Improved Model and the Original YOLOv5s

Model	Class	P/%	R/%	AP/%	mAP/%
YOLOv5s	grave	94.26	92.18	93.57	93.14
	cow	89.35	88.82	89.74	
	car	95.68	94.33	96.10	
Improving YOLOv5	grave	100	100	100	97.25
	cow	91.67	91.76	91.76	
	car	100	100	100	

3.2. Analysis of Model Detection Performance

The confusion matrix for the model on the test set is shown in Figure 2. The classification accuracy for all three target classes exceeds 90%. False positives and false negatives primarily stem from cow targets heavily obscured by vegetation. Overall, both false positive and false negative rates remain at low levels, indicating that the model demonstrates excellent classification and localization performance.

A comparison of the ROC curves for the improved model is shown in Figure 3. The closer the ROC curve is to the upper-right corner of the graph, the better the model's overall performance. As shown in the figure, the ROC curve for the improved model is generally located above and to the right of that of the original model, The area under the PR curve (AP value) for all three target classes is significantly higher than 0.9. Within the range where recall increases from 0 to 0.9, the model's precision consistently remains above 0.9, with only a slight decline when recall approaches 1.0. This demonstrates that the model maintains stable and excellent detection performance across different confidence thresholds. The primary objective of this paper is to achieve accurate identification of the three target classes, and the trained model has met this objective. The specific identification results of the model are shown in Figure 4. The model can accurately identify scattered grave mounds in forested areas, effectively distinguishing them from bare ground and rocks; it can detect herds of cow in forested grasslands, accurately identifying cow in various postures and with partial occlusion; and it can detect vehicles on forest roads, accurately identifying different types of vehicles, thereby meeting the practical needs of large-scale drone patrols in forested areas.

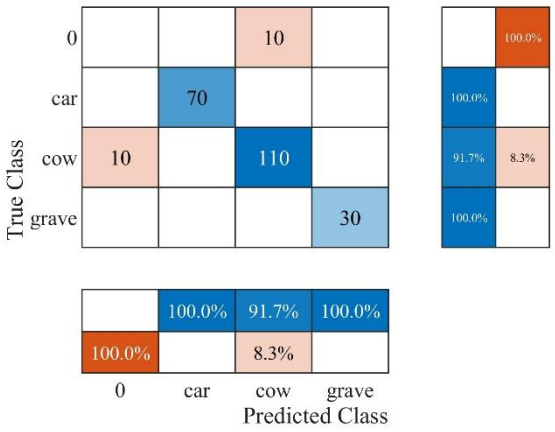

Figure 2: Confusion Matrix

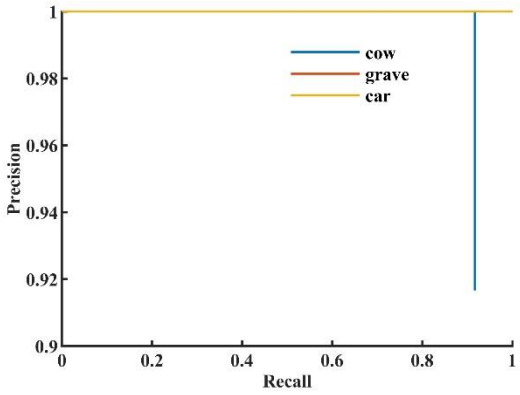

Figure 3: Model Test PR Curve

Figure 4: Model Recognition Performance

4. Discussion

In response to the practical needs of multi-object monitoring in forest areas within the context of smart forestry development, this study has developed a method for the automatic detection of graves, cow, and vehicles based on UAV visible-light imagery and an improved YOLOv5 model. Compared to existing research, the innovations and advantages of this study are primarily reflected in three aspects: A dedicated multi-object detection dataset was constructed for the three core management targets—graves, cow, and vehicles—in forest areas, covering target scenarios with varying vegetation coverage and occlusion levels, thereby providing a data reference for subsequent research on multi-object detection in forestry. To address detection challenges in complex forest backgrounds, the YOLOv5 model was optimized in multiple dimensions: the GhostConv module achieved model lightweighting; the CA attention mechanism enhanced target feature extraction capabilities; the

BiFPN structure improved multi-scale target fusion; and the EIoU loss function optimized bounding box regression accuracy. The improved model achieved a significant increase in detection accuracy and supports real-time detection in the field. This method achieves high-precision multi-object detection using only visible-light imagery from consumer-grade drones. With low data acquisition costs and a low operational threshold, it is easily adoptable by grassroots forestry management departments. It can effectively replace traditional manual patrol methods, significantly improving the efficiency and timeliness of forest area monitoring.

This study still has certain limitations that require further refinement in future research: First, the dataset was primarily collected from the North China larch plantation area in Saihanba. The model's generalization ability across different forest types—such as southern evergreen broadleaf forests and tropical rainforests—and varying topographic conditions remains to be validated. In future work, we will construct multi-regional, multi-temporal datasets to enhance the model's adaptability to different scenarios; Second, the model's detection accuracy for targets heavily obscured by vegetation still has room for improvement. Future work will incorporate Transformer architectures and multi-scale feature enhancement modules to further enhance the detection capabilities for occluded and very small targets; Finally, this study has only achieved image-level object detection. In the future, we will integrate multi-object tracking algorithms, UAV RTK positioning data, and GIS technology to enable spatial localization of objects, trajectory tracking, and spatialized early warnings for fire and grazing risks, thereby establishing a comprehensive "monitoring-early warning-response" operational system.

5. Conclusions

Guided by the practical needs of multi-object monitoring of graves, cow, and vehicles in forested areas, this study developed a method for automatic multi-object detection in forested areas based on UAV visible-light imagery and an improved YOLOv5 deep learning model. The study systematically completed data collection, image preprocessing, dataset construction, model improvement, and performance validation. The main conclusions are as follows: We constructed China's first dedicated dataset for multi-object detection of graves, cow, and vehicles in forest areas. The dataset covers samples of these three target categories across different habitats, scales, and levels of occlusion. Through multi-strategy data augmentation, we effectively enhanced the dataset's diversity, providing a reliable data foundation for deep learning model training. To address the multi-object detection requirements in complex forest backgrounds, the YOLOv5 model was optimized in multiple dimensions for both lightweight performance and accuracy. The improved model achieved an mAP of 97.25% on the test set, representing a 4.11 percentage point increase over the original YOLOv5s model, and is capable of accurately identifying the three target categories against complex forest backgrounds. This study validates the feasibility and superiority of combining consumer-grade drone visible-light imagery with the improved YOLOv5 model for automated multi-object monitoring in forest areas. It effectively addresses the numerous limitations of traditional manual patrol methods and provides efficient, low-cost technical solutions for forest fire hazard identification, grazing ecosystem management, and forest safety supervision, thereby offering significant technical support for the development of smart forestry and digital forestry and grassland management.

Acknowledgements

This research was supported by the Special Science and Technology Project for the Construction of the Chengde National Innovation Demonstration Zone for the Sustainable Development Agenda under Grant 202302F007.

References

[1] Lan Y B, Deng X L, Zeng G L(2019). Advances in diagnosis of crop diseases, pests and weeds by UAV remote sensing. Smart Agriculture,1(02),1-19.

[2] Fan J X, Zhang W H, Zhagn L L, et al.(2023).Vehicle detection method of UAV imagery based on improved YOLOv5.Remote Sensing Information,38(03),114-121.

[3] WU R, LI J, Wang Z,et al.(2025). Vehicle Target Detection System from the Perspective of UAV Based on Improved YOLOv5. Urban Geotechnical Investigation & Surveying,(03),6-11.

[4] Guo X J, Shao Q Q.(2023).Population of kiangs and spatiotemporal variation of its habitat in Sanjiangyuan National Park baded on unmanned aerial vehicle remote sensing.ACTA ECOLOGICA SINICA,43(19),7886-7895.

[5] Wang Y C, Ma J R, Wang J J, et al(2025). Research on lightweight models for detection of large herbivores in the Yellow River Source Region based on UAV Images. PRATACULTURAL SCIENCE,42(06),1538-1551.

[6] Zhang L W, Zhou H, Zhu Q B.(2023). Multi-target tracking of group-housed pigs based on PigsTrack tracker. Transactions of the Chinese Society of Agricultural Engineering,39(16),181-190.

[7] Zhao Y X, Zhang G Q, Li D H, et al.(2025).A high-precision tracking and localization method for monitoring cows. Information and Control,54(01),137-149+160.

[8] Zang K.(2024). Lightweight UAV Remote Sensing Image Vehicle Detection Method Based on Improved YOLOv5s. GEOMATICS & SPATIAL INFORMATION TECHNOLOGY,47(09),86-89.

[9] Xian Y T, Li B, Wan T.(2025). UAV photography detection algorithm based on improved YOLOv5. Computer Measurement & Control,33(04),48-56+66.