Heritage Science

RESEARCH

Open Access

Applying optimized YOLOv8 for heritage conservation: enhanced object detection in Jiangnan traditional private gardens

Chan Gao[1,2], Qingzhu Zhang[2,3,4*], Zheyu Tan[1], Genfeng Zhao[5], Sen Gao[6], Eunyoung Kim[1*] and Tao Shen[7]

Abstract

This study aims to promote the protection and inheritance of cultural heritage in private gardens in the Jiangnan area of China. By establishing a precise visual labeling system and accelerating the construction of a database for private garden features, we deepen the understanding of garden design philosophy. To this end, we propose an improved Jiangnan private garden recognition model based on You Only Look Once (YOLO) v8. This model is particularly suitable for processing garden environments with characteristics such as single or complex structures, rich depth of field, and cluttered targets, effectively enhancing the accuracy and efficiency of object recognition. This design integrates the Diverse Branch Block (DBB), Bidirectional Feature Pyramid Network (BiFPN), and Dynamic Head modules (DyHead) to optimize model accuracy, feature fusion, and object detection representational capability, respectively. The enhancements elevated the model's accuracy by 8.7%, achieving a mean average precision (mAP@0.5) value of 57.1%. A specialized dataset, comprising 4890 images and encapsulating various angles and lighting conditions of Jiangnan private gardens, was constructed to realize this. Following manual annotation and the application of diverse data augmentation strategies, the dataset bolsters the generalization and robustness of the model. Experimental outcomes reveal that, compared to its predecessor, the improved model has witnessed increments of 15.16%, 3.25%, and 11.88% in precision, mAP0.5, and mAP0.5:0.95 metrics, respectively, demonstrating exemplary performance in the accuracy and real-time recognition of garden target elements. This research not only furnishes robust technical support for the digitization and intelligent research of Jiangnan private gardens but also provides a potent methodological reference for object detection and classification research in analogous domains.

Keywords Jiangnan private gardens, Heritage protection, Object detection, Machine learning, YOLOv8

*Correspondence:
Qingzhu Zhang
zhangqz@zjhu.edu.cn
Eunyoung Kim
kim@jaist.ac.jp
[1] Graduate School of Advanced Science and Technology, Japan Advanced Institute of Science and Technology, Nomi 9231292, Japan
[2] Huzhou University, Zhe Jiang Province, Huzhou 313000, China
[3] Northeast Agricultural University, Harbin 150030, Heilongjiang Province, China
[4] Zhejiang Heliang Intelligent Equipment Co. Ltd, Zhe Jiang Province, Huzhou 313000, China
[5] Hangzhou R&F Real Estate Development Co., Ltd, Hangzhou 310000, Zhe Jiang Province, China
[6] Baoding Shengyang Real Estate Development Co., Ltd, Baoding 071000, Hebei Province, China
[7] College of Design and Innovation, Tongji University, Shanghai 200092, China

Introduction

Research background

Jiangnan's private gardens, which are integral to ancient Chinese culture, embody a rich history and are saturated with extensive artistic and cultural connotations [1]. These gardens, representing comprehensive artistic works, mirror the ancient individuals' profound appreciation and pursuit of harmony, balance, and natural beauty [2]. Visitors encounter a blend of art, architecture, horticulture, poetry, and other elements that are meticulously integrated to forge a space that oscillates between tranquility and vitality [3–6].

Designers have invested immeasurable efforts in every detail, employing refined craftsmanship to skillfully organize stones, water, plants, and architecture within confined spaces, thereby creating infinite variation and depth. Every corner, scene, and even each stone carries its own narrative and significance, offering visitors not only diverse sensory stimulations but also a deep comprehension of the philosophical and cultural spirit embedded within these gardens.

The relationship between cultural heritage protection and object detection

The Jiangnan private gardens represent invaluable cultural assets, reflecting the area's profound historical and aesthetic traditions. Protecting these gardens is not just about preserving physical spaces; it's about safeguarding the cultural heritage they embody. However, this task is fraught with challenges, including the absence of standardized records and comprehensive databases. To address these issues, it's essential to integrate object detection technologies into the protection strategies.

Object detection can play a pivotal role in creating visual tags and accelerating the compilation of detailed databases. By identifying and cataloging the various elements within these gardens, object detection helps in building a rich, accessible repository of information. This technology doesn't just aid in documentation; it's a tool for enhancing the protection and promotion of these cultural sites. With precise object detection, it's possible to monitor the condition of various garden elements, detect changes or damages over time, and plan restoration work more effectively.

Moreover, object detection facilitates the creation of interactive digital platforms, transforming the way people engage with cultural heritage. Through virtual tours, educational initiatives, and online exhibitions powered by detailed object recognition, a wider audience can appreciate the beauty and historical significance of the Jiangnan private gardens. This not only broadens the gardens' cultural impact but also fosters a deeper understanding and appreciation among the public, thereby strengthening the case for their preservation.

Motivation and potential advantages of YOLO algorithm in garden heritage protection

The YOLO (You Only Look Once) algorithm is a real-time object detection system. Its primary advantage lies in its ability to detect multiple objects in an image with a single inspection, eliminating the need for multiple scans or sliding window detections. This real-time and accurate nature of the algorithm offers potential benefits for the conservation of private gardens in Suzhou.

Firstly, the elements within Suzhou's private gardens are numerous and often overlap, posing a challenge for traditional object detection algorithms. However, the YOLO algorithm can detect multiple objects in a single glance, effectively managing these complex scenes. With a comprehensive understanding of the entire image, YOLO is adept at handling occlusions and small objects, crucial for the intricate elements and details in Suzhou's private gardens.

Secondly, the real-time capabilities of the YOLO algorithm present potential advantages in garden conservation. For instance, it can be used for real-time monitoring of objects and activities within the garden, allowing for the prompt identification and addressing of behaviors that may damage the garden. Additionally, it can guide the maintenance and repair of the garden in real-time, such as identifying areas that need repair or cleaning. Particularly with the release of YOLOv8 in 2023, which integrates various cutting-edge technologies, the detection accuracy and robustness have been significantly enhanced.

Research objective and paper structure

This study aims to utilize the advanced YOLOv8 to conduct object detection on Jiangnan private gardens, considering the numerous target elements, which often obstruct each other, and the complex, varying-sized backgrounds within the gardens. To overcome these challenges, without the foundation of an open database, we have created a new database containing 4890 images and have categorized the numerous garden elements into four categories: architecture, stone bridges, plants and flowers, and artificial mountains.

To augment the precision and robustness of detection, the following enhancements were made to the initial structure of YOLOv8:

(1) The Diverse Branch Block branch block (DBB) module is added to the backbone layer, replacing the Conv in Bottleneck in C2f, to enhance the model's precision.

(2) A bidirectional feature pyramid network (BiFPN) module is adopted at the neck level to replace the original feature pyramid module, thereby achieving efficient bidirectional cross-scale connections and weighted feature fusion.
(3) The dynamic head (DyHead) module is added to the head layer to enhance the representational capability of the object detection head.

Following these improvements, the results indicate that YOLOv8n-modify improved the accuracy by 8.7% compared to the original YOLOv8, with a mean average precision (mAP) value reaching 57.1%.

The subsequent sections of the paper are organized as follows. Section II introduces the development history of object detection algorithms and the current status and challenges of Jiangnan private garden research. Section III provides a detailed description of the network structure of YOLOv8 and its improved components. Section IV elucidates the collection and processing of the dataset and the parameter settings for model training. Section V presents the results of the experiments. Section VI conducts research discussions, while Section VII explores the prospects and shortcomings of the research.

Materials and methods

Development history of object detection algorithms

The evolution of object detection algorithms has traversed through several pivotal developmental phases, each marked by its unique approaches and challenges.

Knowledge-Based Methods (1970s–1980s): The initial approaches were heavily reliant on rules and heuristic methods, focusing on encoding the shape, color, and texture characteristics of target objects into algorithms [7]. The complexity and instability of the rules, which required substantial human intervention, were the predominant challenges.

Feature-Based Methods (1990s–2000s): This era saw a shift towards utilizing feature descriptors such as SIFT and HOG to represent and detect objects in images, employing sliding- window techniques and classifiers to ascertain the presence of the target object within each window [8, 9]. The main challenges revolved around extracting effective features and ensuring their robustness and discriminability.

Deep Learning-Based Methods (2010s–Present): The advent of deep learning technology, particularly convolutional neural networks, has significantly accelerated advancements in object detection algorithms. Frameworks like R-CNN, Fast R-CNN, Faster R-CNN [10, 11], and YOLO have emerged, capable of performing end-to-end training and prediction on entire images, thereby enhancing the accuracy and efficiency of detection [12–15].

In the era of deep learning, algorithms based on Convolutional Neural Networks (CNN), notably R-CNN, Fast R-CNN, and Faster R-CNN, have gained prominence. These algorithms swiftly generate target areas, utilize CNN to extract features, and subsequently employ a classifier for object recognition. While achieving substantial improvements in object detection accuracy, their computational demands render them unsuitable for real-time applications. Particularly, the YOLO series of algorithms have emerged as prominent in one-stage object detection algorithms, achieving real-time object detection by transforming the object detection task into a dense regression problem. YOLOv8, released in 2023, integrates various cutting-edge technologies, enhancing the accuracy and robustness of detection [16–19].

Improvement of YOLO algorithm in object detect in gardens

In garden research, YOLO algorithm is predominantly applied to monitor plant changes, assisting researchers in understanding ecological shifts within gardens and the potential impacts of these changes on the gardens' ecological environment and cultural value. For instance, Soeb, M. J. A., Jubayer et al. [20] proposed an artificial intelligence-based tea disease detection and identification method based on 4,000 images of different types of tea diseases collected from tea gardens in gardens using the YOLOv7 model. This method is expected to provide strong support for the rapid identification and detection of tea diseases in gardens and reduce economic losses.

Moreover, these technologies have been used in garden management and planning. By automatically detecting and identifying visitor behaviors and distributions, managers can formulate scientifically rational management strategies, aimed at minimizing human interference and ensuring the gardens' original appearance and cultural value are preserved [21–24]. However, most existing research has tended to focus on the detection of specific biological species, while studies on the overall design concept and structure of the gardens are relatively limited.

The design philosophy of Jiangnan's private gardens extends beyond mere object detection. It represents a comprehensive, integrated design system, where each element interconnects with others, forming an inseparable whole [25, 26]. To delve deeper into this design philosophy, a more comprehensive object detection method is required to conduct a thorough study and object recognition of significant garden elements within the private gardens.

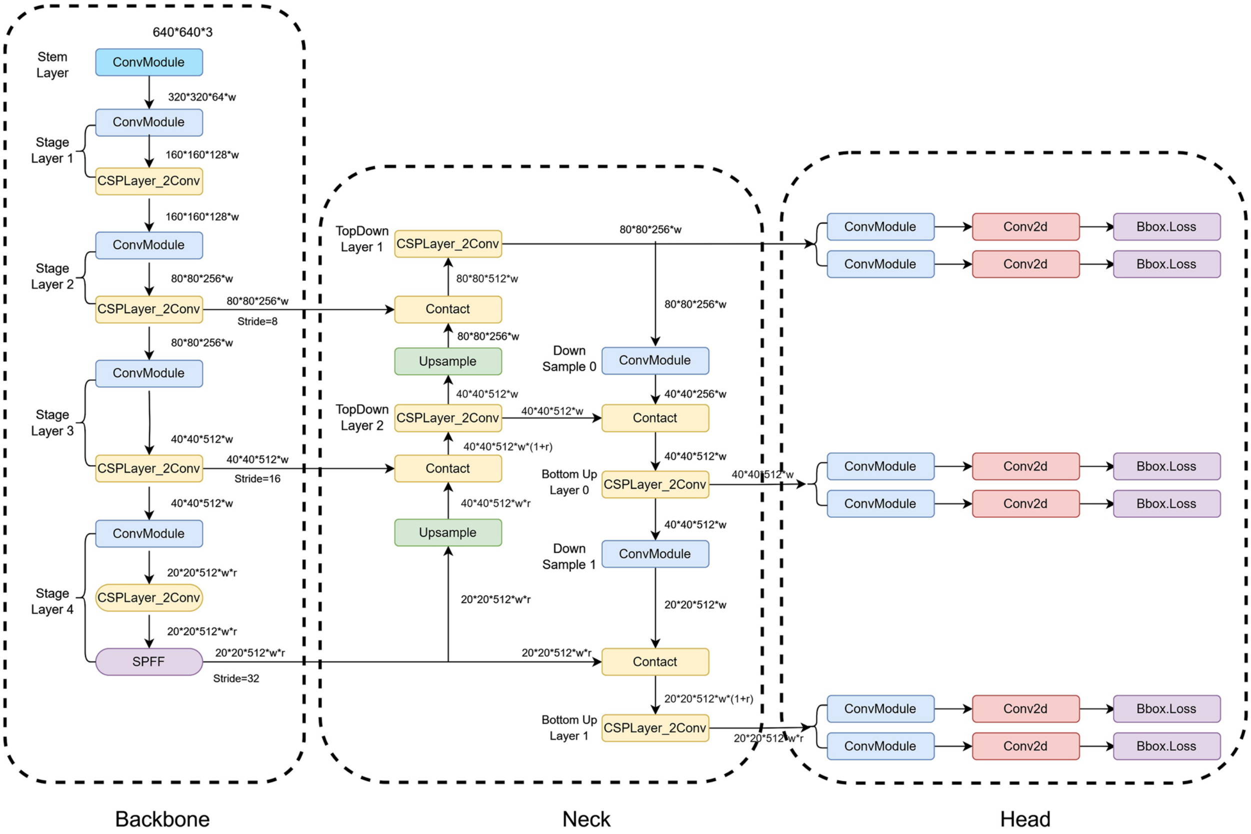


Fig. 1 Detailed structure of YOLOv8

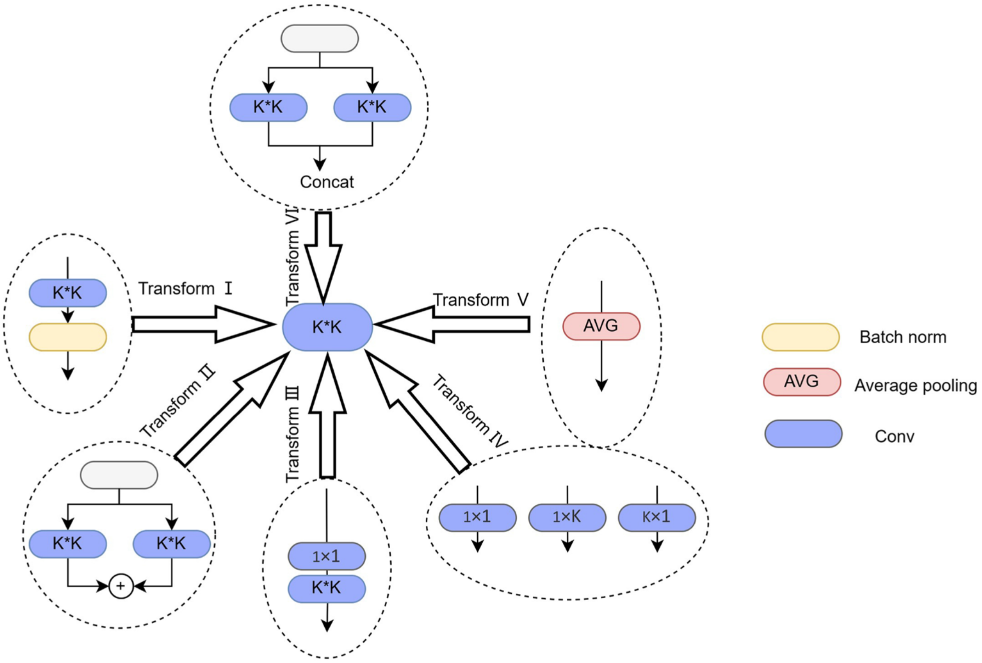


Fig. 2 Six transformations to implement an inference-time DBB by a regular convolutional layer

Pressing issues for Jiangnan private garden object detection based on YOLOv8

(1) Dataset Dilemma: The absence of a public dataset specifically tailored for Jiangnan private gardens necessitates the construction of a large-scale dataset encompassing various landscape elements. Ensuring data accuracy and consistency requires meticulous manual annotation, amplifying the complexity of the task.

(2) Complexity of Object Detection: The multifaceted elements in garden design establish complex spatial relationships, presenting significant challenges for object detection, especially in intricate scenes where numerous targets may be obscured and partially overlapped.

(3) Exploration of Algorithm Performance Improvement: Identifying the optimal algorithm structure for the specific scene of Jiangnan private gardens may necessitate conducting numerous experimental verifications, consuming considerable computational resources and time, given the infancy of this research field and the absence of corresponding research guidance.

YOLOv8 model design and training

YOLOv8 architecture and network structure

YOLOv8, the latest iteration in the YOLO series, employs a network structure that leverages a Feature Pyramid Network and cross-layer connections to seamlessly integrate multi-scale feature information. It amalgamates attention mechanisms and optimization strategies to enhance the accuracy and performance of object detection. The core structure encompasses a backbone network for feature extraction from images, typically utilizing deep convolutional neural network structures such as Darknet or ResNet, and a detection head composed of convolutional and fully connected layers, tasked with predicting the bounding box and class probability of objects.

YOLOv8 approaches the object detection task as a regression problem, utilizing convolutional layers, pooling layers, and fully connected layers to predict object location and class [27–29]. The convolutional layers employ sliding convolutional kernels to extract features from the input data and capture the local spatial structure of the input data. Pooling layers reduce the dimensionality of the feature map, compressing and aggregating features through max-pooling operations, thereby reducing the computation and parameter quantity while enhancing translational invariance. The fully connected layer, positioned at the network's terminus, transforms feature maps into outputs for object detection.

To achieve high-accuracy real-time object detection, YOLOv8 meticulously sets the structure and parameters of the convolutional layers, pooling layers, and fully connected layers, and introduces components such as Anchor Boxes, IoU thresholds, and NMS [30, 31]. Concurrently, it amalgamates various optimization technologies such as data augmentation, batch normalization, and dropout to further enhance performance [32, 33].

Figure 1 illustrates the detailed structure of YOLOv8. It maintains a backbone analogous to YOLOv5 but introduces adjustments on the CSPLayer, incorporating a component termed the C2f module. The C2f module, with two convolutions of the cross-stage partial bottleneck, effectively combines high-level features with contextual information, thereby enhancing detection accuracy.

The first improvement of YOLOv8: integrating DBB module into the backbone layer

The Diverse Branch Block (DBB) adopts an innovative design, replacing Conv in the Bottleneck of C2f with DBB, introducing a multi-branch structure with different receptive fields and complexities, significantly enhancing the detection accuracy of the original model. The DBB

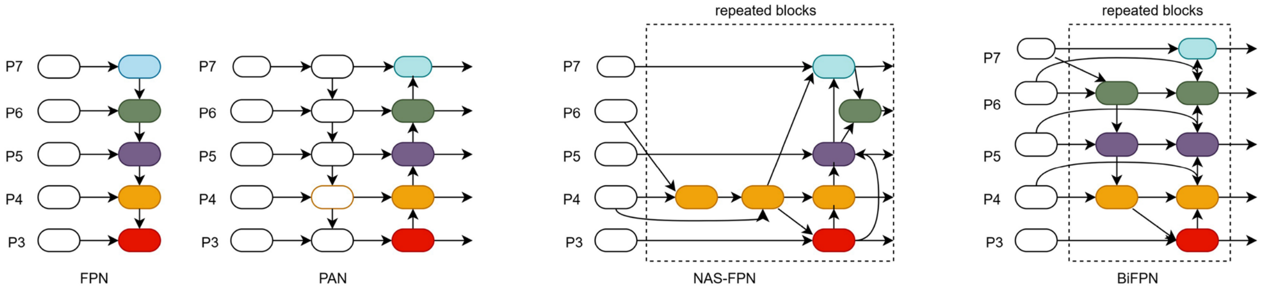


Fig. 3 Bidirectional feature pyramid network (BiFPN) is superior to other networks

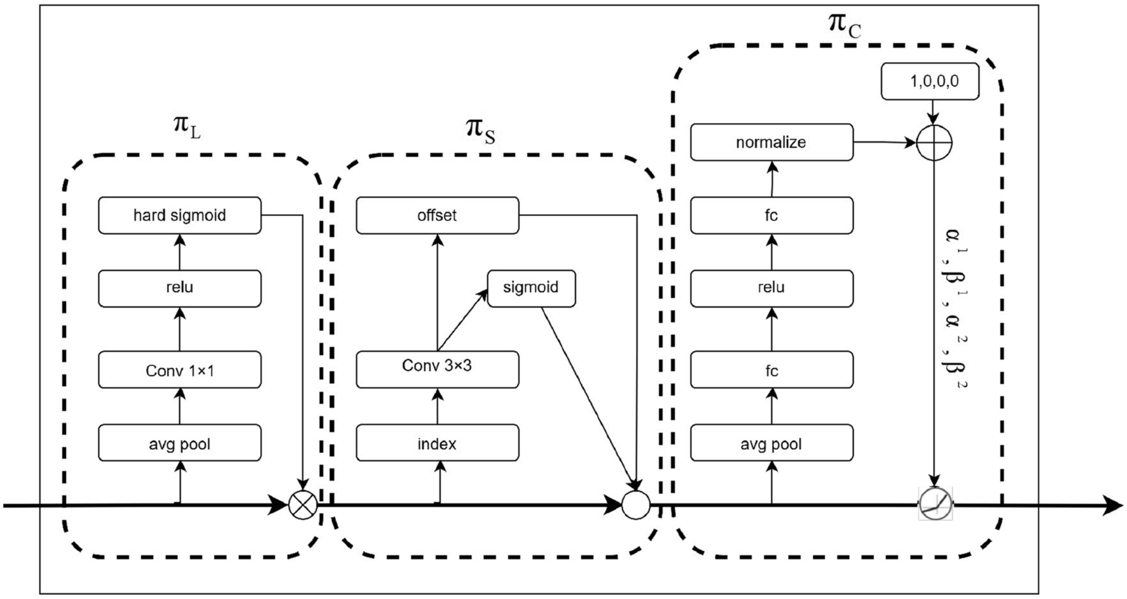


Fig. 4 Dynamic head (Dyhead) framework

design, inspired by the inception architecture, combines multi-scale convolution, sequential 1×1—K×K convolution, average pooling, and branch addition multi-branch topology. This structure can enrich the feature space, provide different complexity receptive fields and paths, and enhance the feature-extraction capability of the model.

The DBB design not only enhances the model's accuracy during training but can also be equivalently converted into a single convolution operation during the inference phase. This means that we can use the DBB to replace any K×K convolution in the model during the training phase, and in practical applications, the DBB can be converted back into a K×K convolution through six structural reparameterization conversion methods (Fig. 2), thereby improving the detection accuracy without increasing the model complexity and computational quantity and without worrying about additional inference time costs.

The second improvement of YOLOv8: integrating BiFPN into the neck layer

A bidirectional feature pyramidBi-directional Feature Pyramid Network (BiFPN) is employed to replace the original feature pyramid module, aiming to fuse features more efficiently. The BiFPN is an optimized cross-scale connection method designed specifically to enhance the efficiency and accuracy of feature fusion networks.

Compared to traditional feature pyramids, BiFPN has several key optimizations (Fig. 3):

(1) Simplifying the network structure: By removing nodes with only a single input, unnecessary feature transmission is reduced, thereby simplifying the network.
(2) Enhancing feature fusion: When the original input and output nodes are at the same level, the BiFPN adds additional connections to more thoroughly fuse feature information without adding extra computational costs.
(3) Reusing feature network layers: A BiFPN views each bidirectional path as a separate feature network layer and repeats this layer multiple times to achieve deep feature fusion. This design provides more flexibility than traditional top-down and bottom-up paths and can adapt to resource limitations at different numbers of layers.

Through these optimizations, the BiFPN not only significantly improves the model performance for object detection and semantic segmentation tasks, but also ensures an enhancement in accuracy.

The third improvement of YOLOv8: integrating an attention mechanism-based object detection head, DyHead, into the head layer

We adopted a new dynamic head framework (Dyhead) for the object detection head and attention mechanism. This method coherently combines various self-attention mechanisms across scale-aware feature levels, spatially aware positions, and task-aware output channels, thereby significantly enhancing the representational capability of

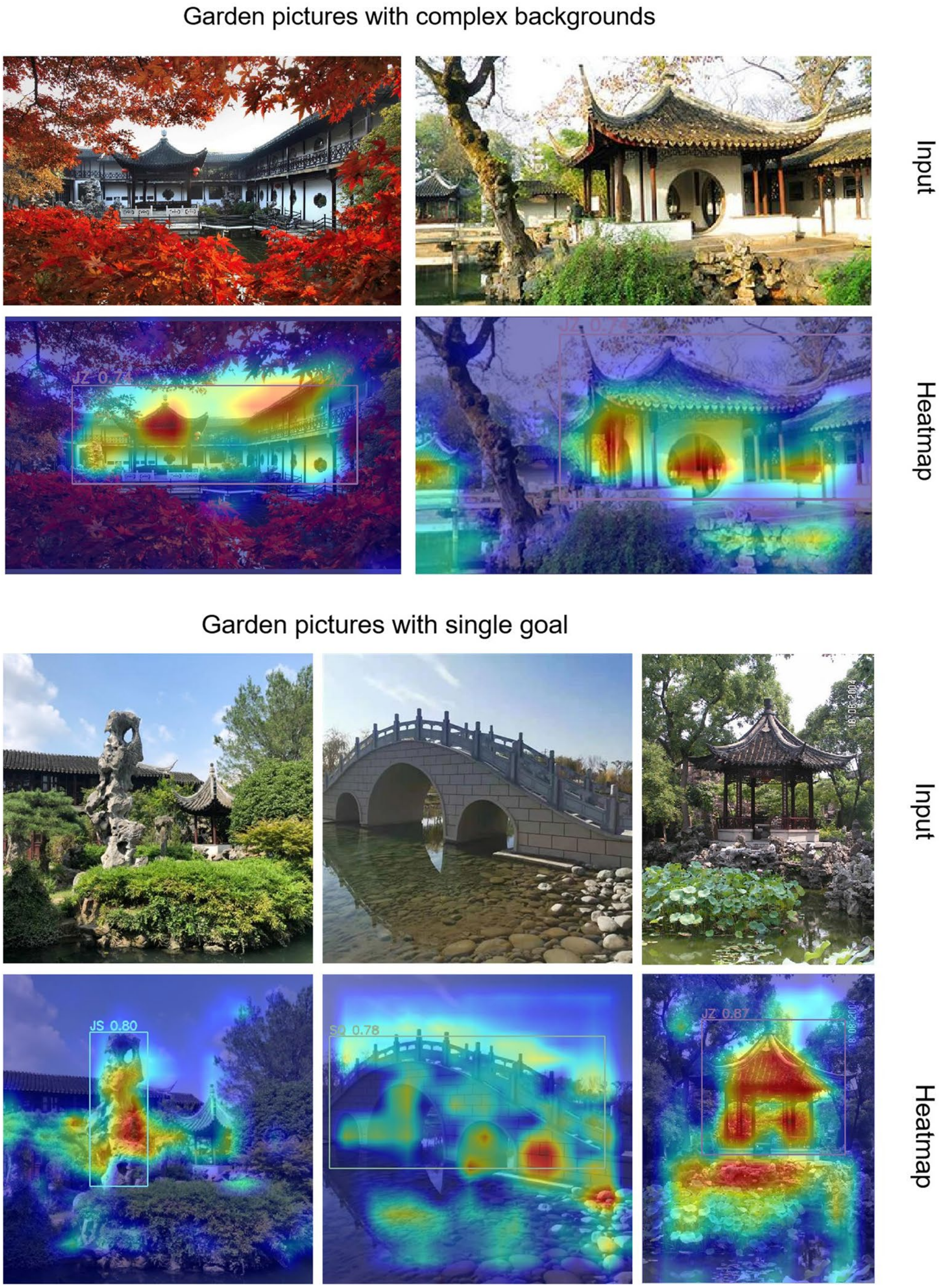

Garden pictures with complex backgrounds
Input
Heatmap
JZ 0.74
JZ 0.74
Garden pictures with single goal
Input
Heatmap
JS 0.80
SQ 0.78
JZ 0.87

Fig. 5 Attention heat map after applying the DyHead mechanism

the object detection head. Specifically, this model stacks scale attention, spatial attention, and task attention.

π_L represents scale-aware attention. π_S represents spatial attention, employing Deformable Convolution V2 (including offset and feature amplitude modulation). π_C represents channel attention and channel modeling through two fully connected neural networks (Fig. 4).

After applying the DyHead mechanism, we obtain a pair of direction-aware feature maps. These maps can act complementarily on the input features, thereby enhancing the object representation. As shown in

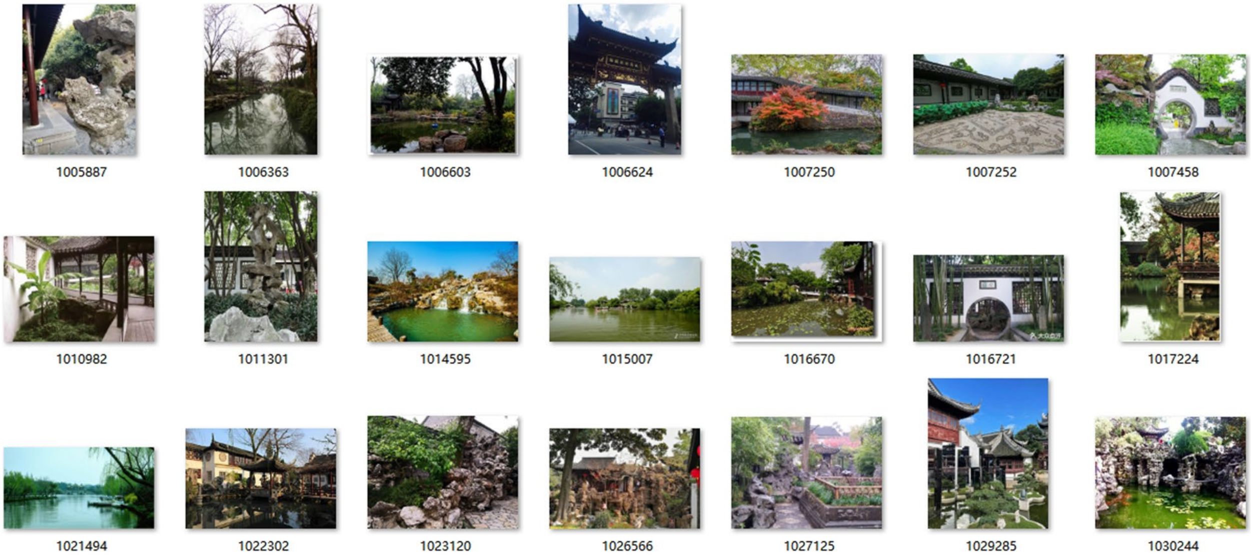


Fig. 6 Part of the database image

Fig. 5, the top is the input image, while the bottom is the heatmap after attention mechanism processing. The below image shows a pure background, whereas the above image shows a complex background. The red part of the heatmap represents the information that the neural network considers to be more important, whereas the blue part represents background information unrelated to classification.

Data collection and annotation

Classification of garden elements

To facilitate the application of object detection algorithms in recognizing and classifying garden elements, a preliminary definition of garden element classification is essential. Traditional garden layouts and designs are notably complex, especially in the realm of garden architectures, which manifest in a myriad of forms. For instance, pavilions, a specific type of small garden architecture, can be classified from various perspectives, such as plans, roofs, and walls, owing to their diverse structures and styles [19]. However, a detailed classification of each form of garden architecture may result in insufficient sample quantities to support learning by the object detection algorithm. Consequently, a classification method that can encompass various garden architectures while ensuring a sufficient sample size, is required.

Based on the study of Jiangnan private gardens, they are categorized into four major classes:

(1) Architecture is marked as JZ: encompassing halls, palaces, and other structures, primarily utilized for dwellings and entertaining guests.
(2) Stone bridges are marked as SQ: manifesting in various forms such as curved, straight, crescent, moon-viewing, and dragon-crossing.
(3) Rockeries are marked as JS: subdivided into peak rock, waterstone, cave, and peculiar stone styles.
(4) Plants were marked as ZW: including bonsais, flowers, vines, grasslands, and aquatic plants.

Data collection and preprocessing

To facilitate the study of the Jiangnan private gardens, a specialized large-scale dataset must be constructed, given the absence of a public dataset for Jiangnan private gardens. Initially, 3280 representative images of the Jiangnan private gardens were sourced from the internet, covering gardens such as the "Four Famous Gardens": Nanjing Zhan Garden, Suzhou Liu Garden, Suzhou Zhuozheng Garden, and Wuxi Jichang Garden, and other gardens such as Shanghai Yu Garden, Yangzhou Slender West Lake, Ge Garden, He Garden, Suzhou Canglang Pavilion, Lion Forest Garden, and Nantong Shuihui Garden, etc. Additionally, 1610 photos were personally captured in Jiangnan private gardens, constructing a database containing 4890 images (Fig. 6). These images span four categories; to enhance the robustness of the model, 1% of background images were also added, ensuring the images display various angles and different lighting conditions of the gardens.

Dataset annotation

Accurate dataset annotation is critical for successful object detection. Rectangular boxes were utilized to annotate each object, and the corresponding class labels were assigned to each box. The architecture was labeled JZ, stone bridges SQ, rockeries JS, and plants ZW. The LabelImg tool was used for manual annotation to enhance the efficiency and accuracy of annotation. During annotation, particular attention was paid to ensure that each object's bounding box was tight, the issues of vegetation occlusion and architectural segmentation were addressed, and buildings of multiple categories were annotated separately. For each image, corresponding annotation files were generated, such as YOLO format label files (each line contains a target's class and bounding box coordinate information) or COCO format label files (each file contains all the target information for one image).

Data Preprocessing and Augmentation Methods

The dataset was randomly divided using the holdout method, with a split ratio of test set: training set: validation set of 80%: 10%: 10%. Ultimately, 3912 training images, 489 test images, and 489 validation images were obtained. First, the size of the images was adjusted to dimensions like 640 × 640, and the pixel values were normalized between 0 and 1 to enhance the training effect and stability of the model. Second, data augmentation was performed using several methods to enrich the dataset.

(1) A total of 35% of the images were mirror-flipped to increase image diversity and reduce the dependence of the model on mirror symmetry.
(2) 25% of the images were randomly flipped to enhance the ability of the model to recognize objects at different angles.
(3) 20% of the images were randomly grayscaled to simulate different lighting conditions, thereby making the model more robust.
(4) A 15% random Gaussian noise was added to the images to help the model adapt to the low resolution appearing in the old garden photos.
(5) 35% of the images were randomly occluded, and objects were separated to help the model better handle occluded and overlapping objects in the Jiangnan private gardens.

Maintaining the consistency and accuracy of the annotation information is crucial during data preprocessing and augmentation operations. After size adjustment, cropping, or augmentation, the position and class information of the bounding boxes remained consistent with those of the original images before processing.

Computer configuration and parameter settings

In our experimental environment, we used the YOLOv8 model developed by Ultralytics, version 0.114. The experimental environment was configured using Python 3.9.0, VScode (1.76.0) IDE, and CUDA 10.2. All model

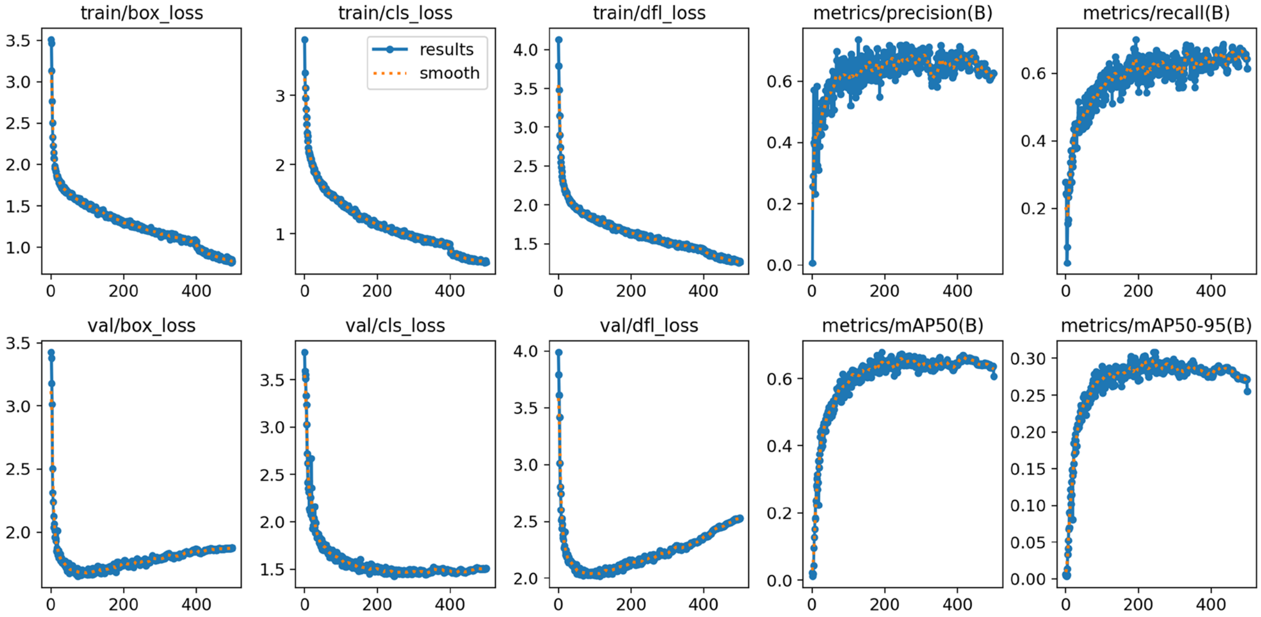


Fig. 7 Performance parameters of model training for 500 rounds

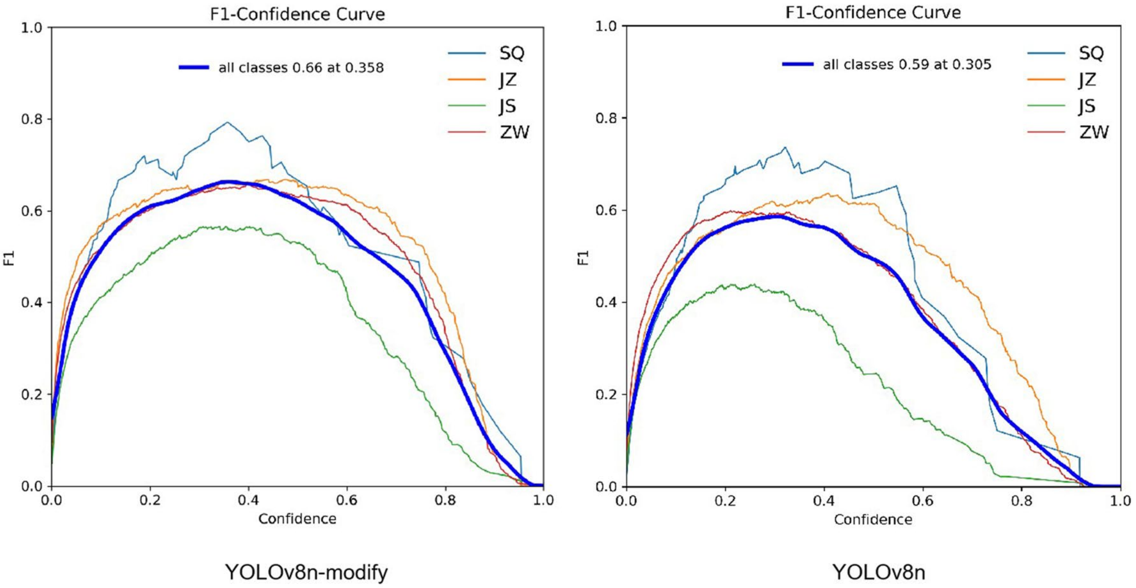


Fig. 8 F1 and confidence curve

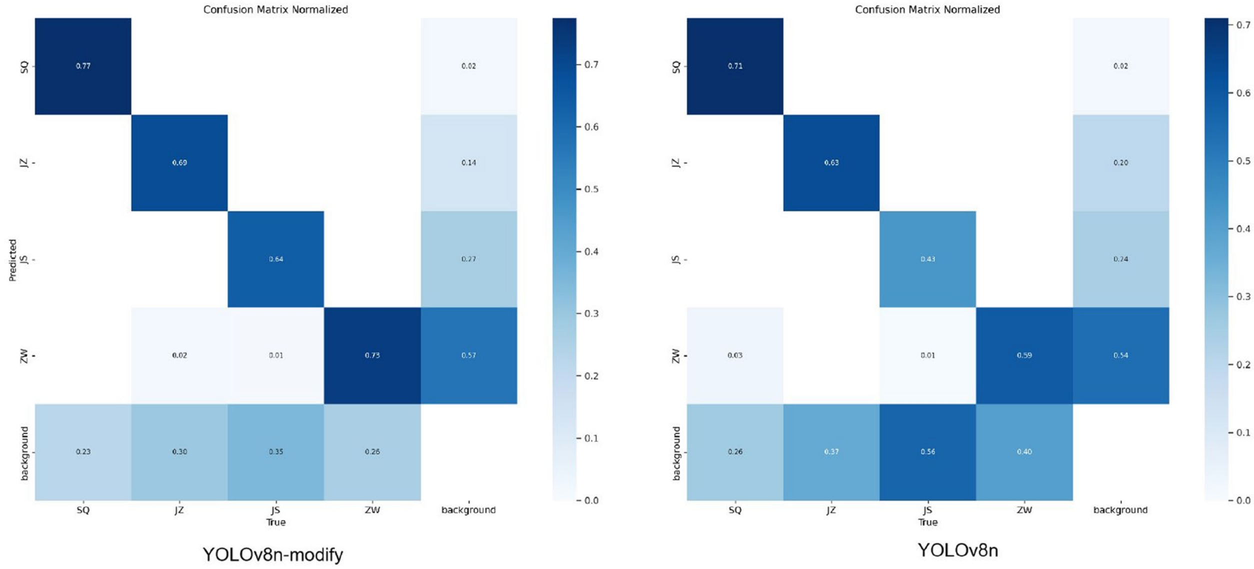


Fig. 9 Confusion matrix

training and testing were executed on an NVIDIA TITAN V (12 GB). During the training phase, the optimizer choice was left adaptive, automatically selecting between SGD and Adam and adapting to the gradient descent characteristics of different tasks. The SGD initial learning rate was set to 1E−2, the Adam initial learning rate to 1E−3, and the weight decay to 5E−4. The weight decay and momentum factors were set to 0.0005 and 0.937, respectively. In addition, three warm-up stages of 0.8 momentum were conducted, and the cosine annealing method was utilized to decay the learning rate, with 500 epochs per experiment and a batch size of 16. The entire model-training process spanned approximately 6.5 h. In this paper, the YOLOv8 model and all comparative models adopt an 'N' model size, ensuring fairness in the evaluation. The YOLOv8n model, compared to other sizes, offers a balanced compromise between speed and accuracy, providing optimal performance for real-time

Table 1 The ablation experiment results

Modules	Add DBB	Add BiFPN block	Add DyHead attention	Precision (%)	Recall (%)	mAP@0.5(%)	mAP@0.5:0.95(%)
YOLOv8n	×	×	×	57.4	53.5	55.3	26.1
YOLOv8n-C2f-DBB	○	×	×	64.2	51.2	58.2	28.1
YOLOv8n-bifpn	×	○	×	60.2	51.4	56.8	27.5
YOLOv8n-dyhead	×	×	○	56.8	48.5	51.7	26.9
YOLOv8n-bifpn-dyhead	×	○	○	61.1	52.7	60.6	28.9
YOLOv8n-C2f-DBB-bifpn	○	○	×	64.8	53.2	62.5	30.1
YOLOv8n-C2f-DBB-dyhead	○	×	○	63.2	52.0	58.8	29.8
YOLOv8n-modify	○	○	○	66.1	46.7	57.1	29.2

applications without excessively taxing computational resources.

In summary, the data collection, annotation, and preprocessing phases are pivotal for ensuring the robustness and accuracy of the object detection model. A detailed classification of garden elements, meticulous annotation, and strategic data augmentation methods were employed to enhance the learning capability and generalization of the model. Furthermore, careful selection of computer configurations and parameter settings during the training phase is crucial to facilitate efficient learning and optimization of the model, ensuring that it can accurately detect and classify objects in the Jiangnan private gardens.

Evaluation and analysis of model performance

Metrics for evaluation

To verify the efficacy of the modified YOLOv8n model, herein referred to as YOLOv8n-modify, using our curated dataset, we used four prevalent metrics: Precision, Recall, F1 Score, and mAP, complemented by the confusion matrix to gauge the model's performance. An Intersection Over Union (IOU) threshold of 0.7 and a confidence threshold of 0.25 were established to impartially assess experimental outcomes.

Outcomes of model training

Precision exhibited an expeditious learning trajectory during the initial phases of 500 training iterations, encountered periodic fluctuations in precision, and attained a semblance of mid-term stability. This suggests that the model navigates through various data feature challenges during the learning trajectory and identifies local optimal solutions at certain junctures. Despite experiencing some precision fluctuations in subsequent stages, it was predominantly sustained within a higher range (0.69666–0.73635), signifying that YOLOv8n-modify has assimilated the principal information of garden targets (Fig. 7).

The recall rate (Fig. 7) went through a process of fluctuation. After the initial value of 0.27786 gradually decreases to 0.03762, the performance of the model begins to gradually increase and shows an upward trend in several stages until it reaches 0.70001.

The YOLOv8n-modify model exhibits a commendable performance enhancement during training, as evidenced by the consistent decline in all (Fig. 7) three loss components. The box_loss, representing bounding box accuracy, markedly decreases from 3.4259 to around 1.65, indicating a significant improvement in the model's ability to locate objects precisely within the image. Similarly, the cls_loss, reflecting the object classification accuracy, also shows a substantial decline from 3.7899 to approximately 1.43, demonstrating the model's increasing proficiency in correctly identifying object classes. The dfl_loss, denoting distribution fitting or another complex aspect of the model, though showing greater variance, trends downward from 3.9893 to about 2.02, suggesting the model's

Table 2 YOLOv8n-modify comparative analysis with alternative methods

Modules	Adopted modules	Precision (%)	Recall(%)	mAP@0.5(%)	mAP@0.5:0.95(%)	FPS(%)
YOLOv3n	None	55	48.6	53.1	25.8	27.2
YOLOv5n	None	57.4	51	51	22.4	38.3
YOLOv8n	None	57.4	53.5	55.3	26.1	38.9
YOLOv8n-modify	DBB + BiFPN + DyHead Attention	66.1	46.7	57.1	29.2	14.8

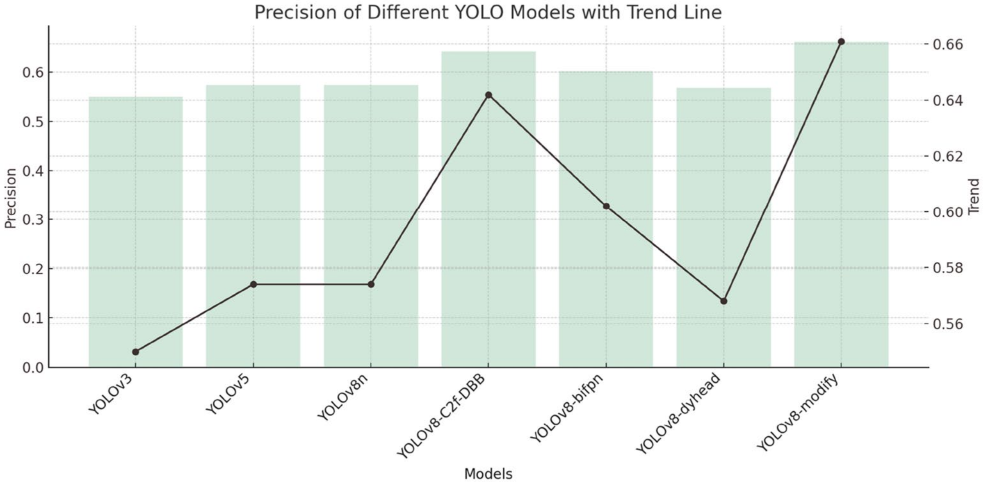


Fig. 10 Comparison of precision between YOLOv8n-modify and other models

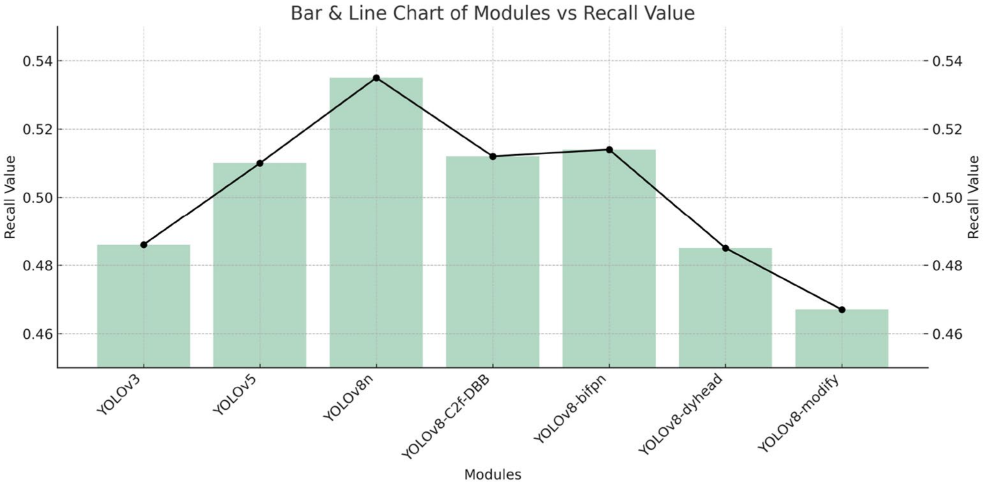


Fig. 11 Comparison of recall value between YOLOv8n-modify and other models

enhanced capability in handling complex data distributions or refining predictions. These trends collectively underscore the model's robust learning capacity.

The F1 versus confidence curve (Fig. 8) shows the F1 score of the model at different confidence thresholds. Experiments show that, compared with YOLOv8, the F1 value of the YOLOv8n-modify curve is higher than that of the original model at most confidence thresholds, indicating that the improved model performs better at various confidence levels. Moreover, the curve of the improved model was smoother than that of the original model, indicating that the performance of YOLOv8n-modify was more stable at different confidence levels. The F1 value reaches 0.66 when the confidence threshold is 0.358, which is 7% higher than YOLOv8.The value range of mAP@0.5 increases from 0.01213 to 0.67807.

In the confusion matrix (Fig. 9), the left picture is YOLOv8n-modify, and the right picture is the YOLOv8n initial result. In YOLOv8-modify, the recognition accuracy rates of the SQ, JZ, JS, and ZW categories are 0.77, 0.69, 0.64, and 0.73, respectively. It is 0.06, 0.08, 0.21, and 0.14 higher than the recognition accuracy of the SQ, JZ, JS, and ZW categories in YOLOv8 (0.71, 0.61, 0.43, and 0.59), respectively. Notably, the high accuracy emphasizes the robustness of the model in identifying these categories. However, the model occasionally misclassified backgrounds as ZW classes. This phenomenon may be attributed to overlapping features or shared properties between background and ZW categories.

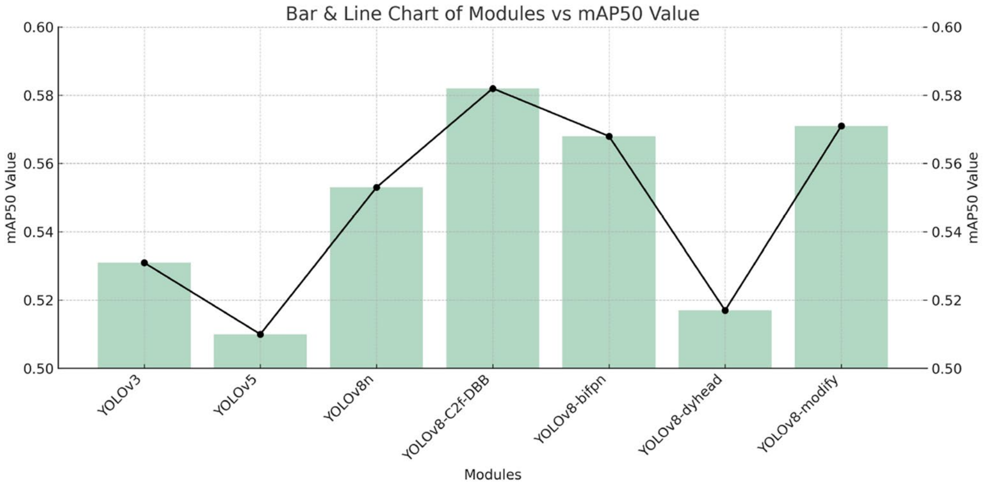


Fig. 12 Comparison of mAP50 between YOLOv8n-modify and other models

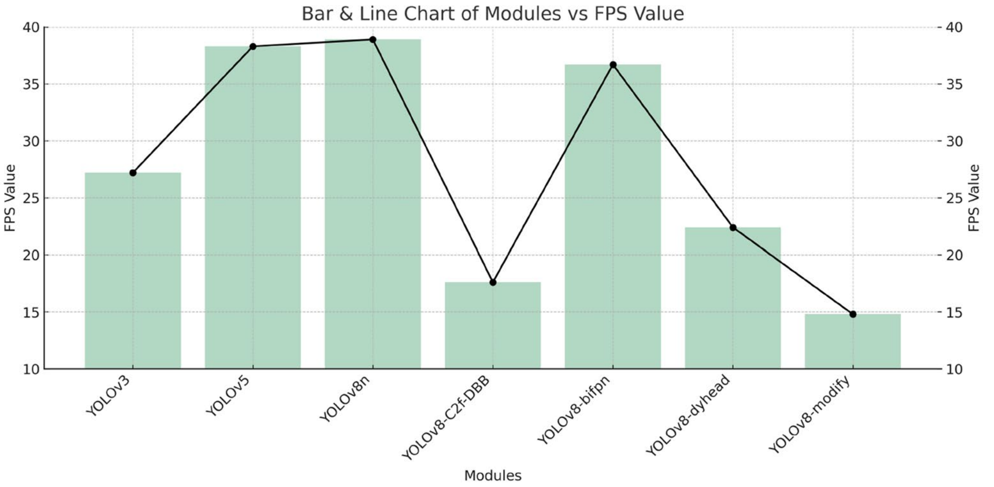


Fig. 13 Comparison of FPS between YOLOv8n-modify and other models

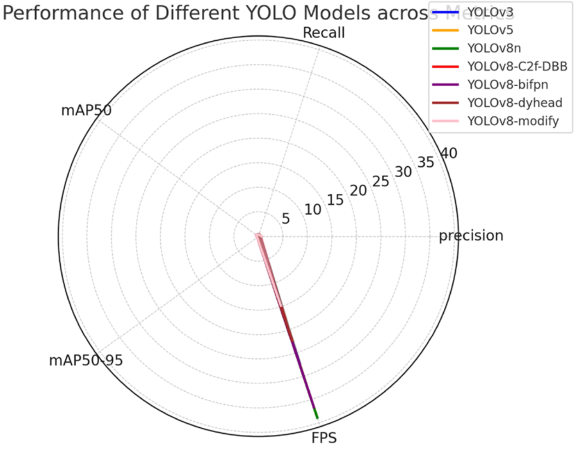


Fig. 14 Radar chart

Ablation study

To substantiate the optimization impact of the three enhancement strategies on the garden dataset, ablation studies were conducted to assess the efficacy of each enhancement strategy. The experimental results are presented in Table 1.

Upon integrating the three modules into the network model for the ablation experiments, DBB enhanced the detection accuracy without increasing the model complexity and computational burden. The BiFPN efficiently fuses features, whereas DyHead overlays scale, spatial, and task attention, thereby fortifying the model's feature extraction capabilities. Table 1 displays the results of all tests, wherein each model, set with consistent hyperparameters and pretraining weights, was trained for 500

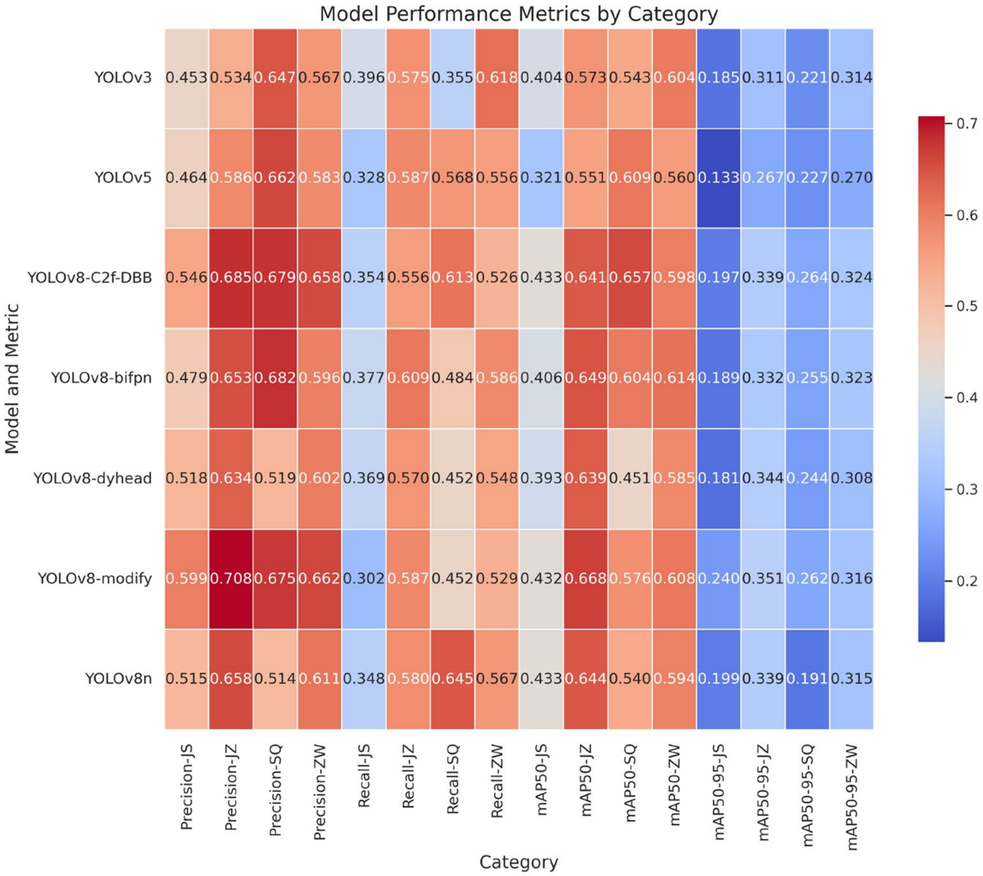


Fig. 15 Heatmap

epochs. The findings indicated that the application of the three module modifications to the optimized YOLOv8 achieved a precision of 66.1%, recall of 46.7%, and mAP50 of 57.1%.

Comparative analysis with alternative methods

In the final analysis, the enhanced detection method was juxtaposed with other detection methodologies, namely YOLOv3, YOLOv5, YOLOv8, in addition to YOLOv8-modify, as depicted in the figure. To ensure a fair comparative analysis, the operational environment and network parameters were maintained consistently, and all methodologies were trained until convergence to realize optimal performance. Table 2 presents the detection outcomes of each method evaluated using an identical test set.

The results indicate that the proposed method is superior to other existing methods for object detection. The results demonstrated that the application of modifications to the three modules in the optimized YOLOv8 achieved a precision of 66.1%, a recall of 46.7%, and an mAP50 of 57.1%. This significant enhancement in accuracy validates the efficacy of the proposed method for object detection tasks. Compared with other models, the improved model is more suitable for detecting objects in the complex scenes of Jiangnan private gardens, and the enhancements are effective.

Figure 10 shows the performance of the different versions of the YOLO model in terms of precision. The light green bars represent the accuracy of each model, and the line chart shows the trend in the accuracy. The YOLOv8n-modify model has excellent performance in terms of accuracy, reaching 0.661, whereas the accuracies of the other models are between 0.55 and 0.65.

Figure 11 illustrates that the recall rates for all models range between 0.467 and 0.535. YOLOv8n achieves a peak recall of 0.535 without incorporating additional modules. Conversely, the introduction of multiple modules to YOLOv8 (denoted as YOLOv8-modify) resulted

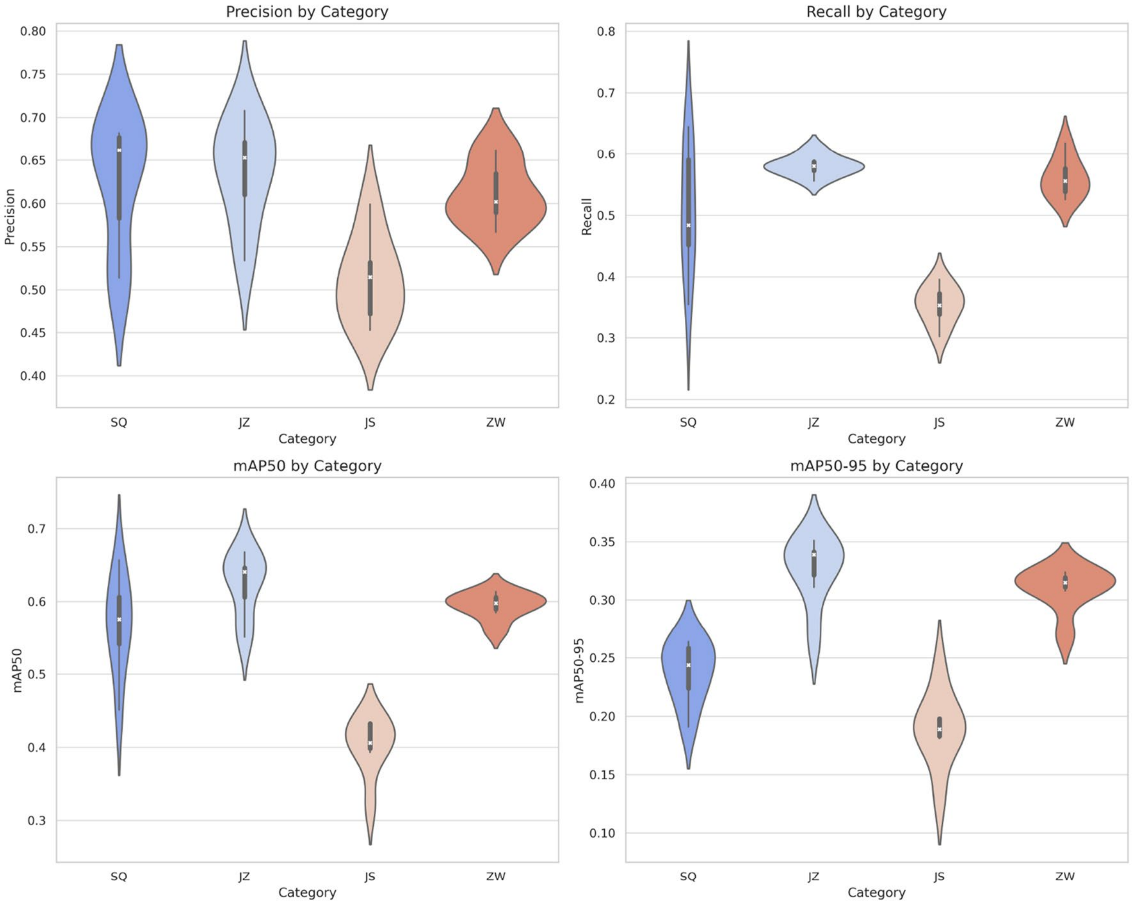


Fig. 16 Violin plot

in a reduced recall of 0.467, indicating an increase in detection accuracy while potentially overlooking certain detectable objects. This phenomenon is partially attributed to the actual image content because numerous images in the database are densely populated with greenery, buildings, and various visually complex detection objects. Some missed detections were reasonable, since our algorithm was initially configured to prioritize high detection accuracy.

mAP50 represents the average accuracy when the IoU (Intersection over Union) threshold is 0.5. As shown in Fig. 12, YOLOv8-C2f-DBB showed the highest mAP50 value (0.582), followed by the YOLOv8n-modify (0.571).

FPS, which denotes the number of frames processed by the model per second, correlates with the real-time performance of the model. As shown in Fig. 13, YOLOv8n had the highest FPS of 38.9. Conversely, YOLOv8n-modify exhibited the lowest FPS value among all models, registering 14.8. This implies that although YOLOv8n-modify demonstrates commendable accuracy, its real-time performance is suboptimal.

The comprehensive algorithm evaluation of YOLOv8n-modify

The radar chart (Fig. 14) encompasses five indicators: precision, recall, mAP50, mAP50-95, and FPS (frames per second). Observing the shape of each color enables an understanding of the performance of the model across each indicator. The chart reveals that YOLOv8n-modify (depicted in pink) outperforms in terms of precision, mAP50, and mAP50-95, albeit slightly underperforming in FPS. Conversely, YOLOv8n and YOLOv5 exhibited commendable performance in FPS but slightly lagged in other indicators.

The heat map in Fig. 15 illustrates the performance of various YOLO models across four indicators (Precision, Recall, mAP50, and mAP50-95) in disparate categories. The color depth signifies the magnitude of the value, where darker colors denote lower values, and lighter colors denote higher values. YOLOv8n-modify exhibits proficient performance in most categories, especially in the precision and mAP50-95 of the JZ category. Compared to other models, YOLOv5 and

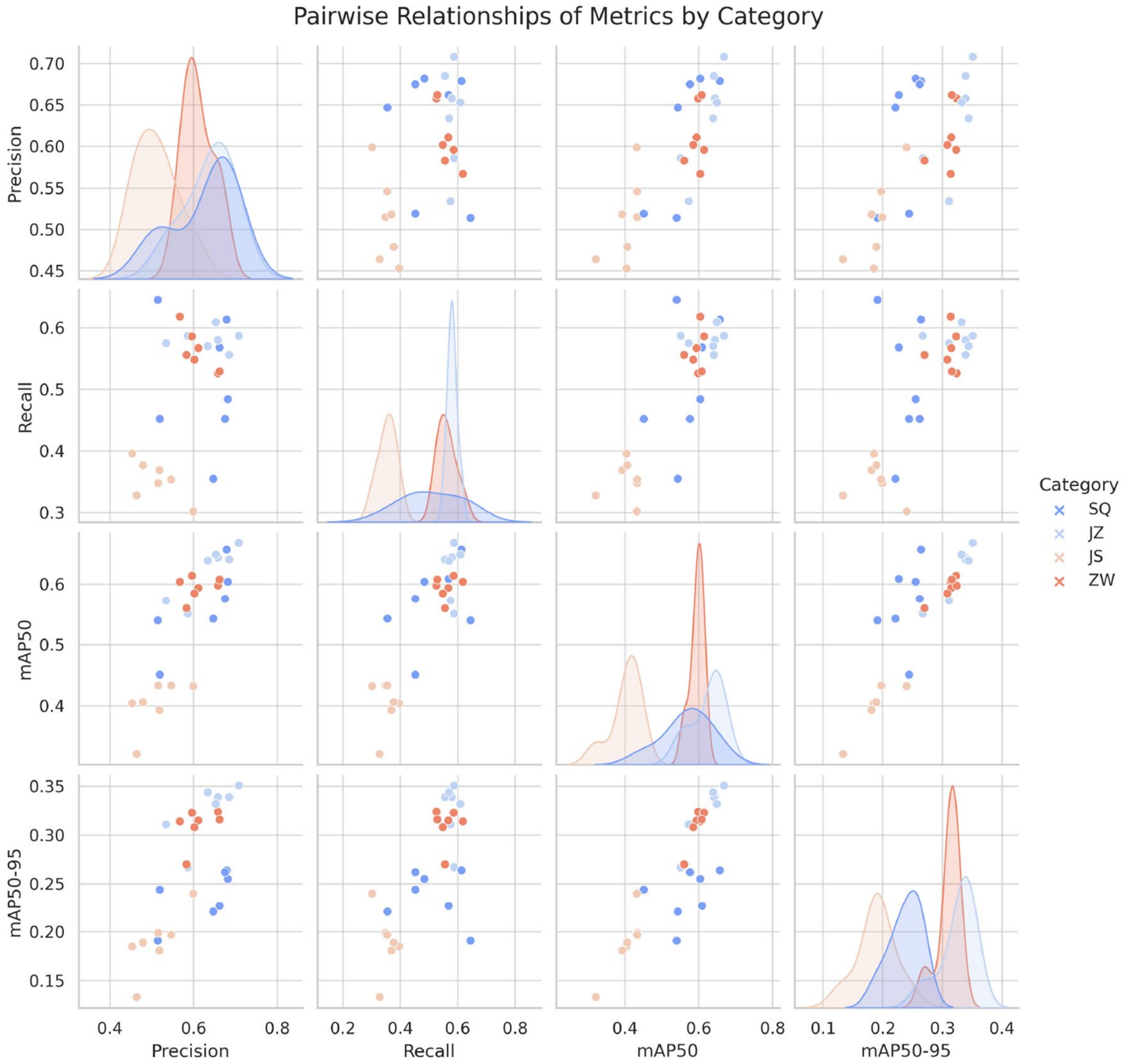


Fig. 17 Pair plot

YOLOv8-C2f-DBB also exhibit proficient performance in some categories but do not parallel YOLOv8-modify's overall performance. This indicates that the version and modifications of the algorithm significantly influence performance.

The violin plot in Fig. 16 illustrates the distribution of each category across the four indicators. In the Precision of the SQ category, most model scores oscillate between 0.5 and 0.7, with YOLOv8n-modify reaching 0.675, exhibiting the highest density and performance. Concurrently, for the Recall of the ZW category, most model scores fluctuate between 0.5 and 0.6, but YOLOv3 slightly leads with a score of 0.618. Additionally, for the mAP50 of the JS category, most model scores fluctuate between 0.4 and 0.6, but YOLOv8n-modify reaches 0.599, surpassing other models. This chart provides an intuitive understanding of the performance disparities of the different models across each indicator and the models that excel in specific categories.

The pair plot in Fig. 17 illustrates the pairwise relationships between the four indicators. There is a positive correlation between Precision and Recall in the JZ and ZW categories. When the Precision in the JZ category reaches 0.7, the recall is approximately 0.6. Moreover, in all categories, mAP50 and mAP50-95 exhibited a strong positive correlation, especially in the SQ and ZW categories, where mAP50 was 0.6 and mAP50-95 was 0.3. This indicates that these two indicators are similar across all categories.

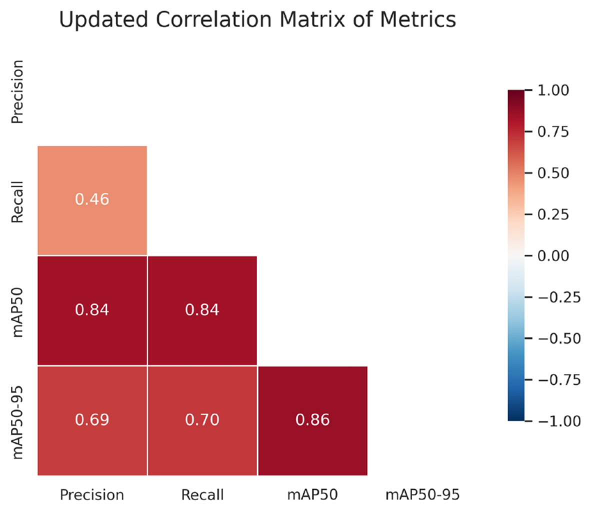


Fig. 18 Correlation matrix

The correlation matrix (Fig. 18) provides a quantitative measure of the correlations between the four indicators. From the chart, it is evident that the correlation between mAP50 and mAP50-95 is exceedingly high, approaching 1, indicating that when one indicator increases, the other also increases. However, the correlation between Precision and Recall was low, indicating that in some instances, enhancing precision may necessitate sacrificing recall.

Discussion

Interpretation and discussion of experimental results

(1) Integration of DBB into the backbone layer

The incorporation of the DBB has significantly enhanced the accuracy of object detection in the model. The DBB enriches the feature space via a multi-branch structure, bolstering the model's detection accuracy without augmenting the computational complexity during the inference phase. It employs multiscale convolution and sequential convolution technologies to extract multiscale features and optimize the receptive field, maintaining efficient computational performance during the inference phase by converting it into a single convolution operation.

(2) Implementation of BiFPN in the neck layer

BiFPN amplifies the efficiency and accuracy of a feature fusion network by optimizing cross-scale connections. It streamlines the network structure, minimizes unnecessary feature transmission, fortifies feature fusion, and deepens feature fusion by iteratively repeating the feature network layer, thereby adding additional connections without escalating the computational costs.

(3) YOLOv8 enhancement using DyHead

YOLOv8 augments the representational capability of object detection by introducing DyHead based on attention mechanisms in the head layer. DyHead concentrated on scale-aware, spatial-aware, and task-aware self-attention mechanisms and overlaid scale attention, spatial attention, and task attention facilitate effective information exchange between different feature levels and spatial positions, thereby optimizing the detection performance.

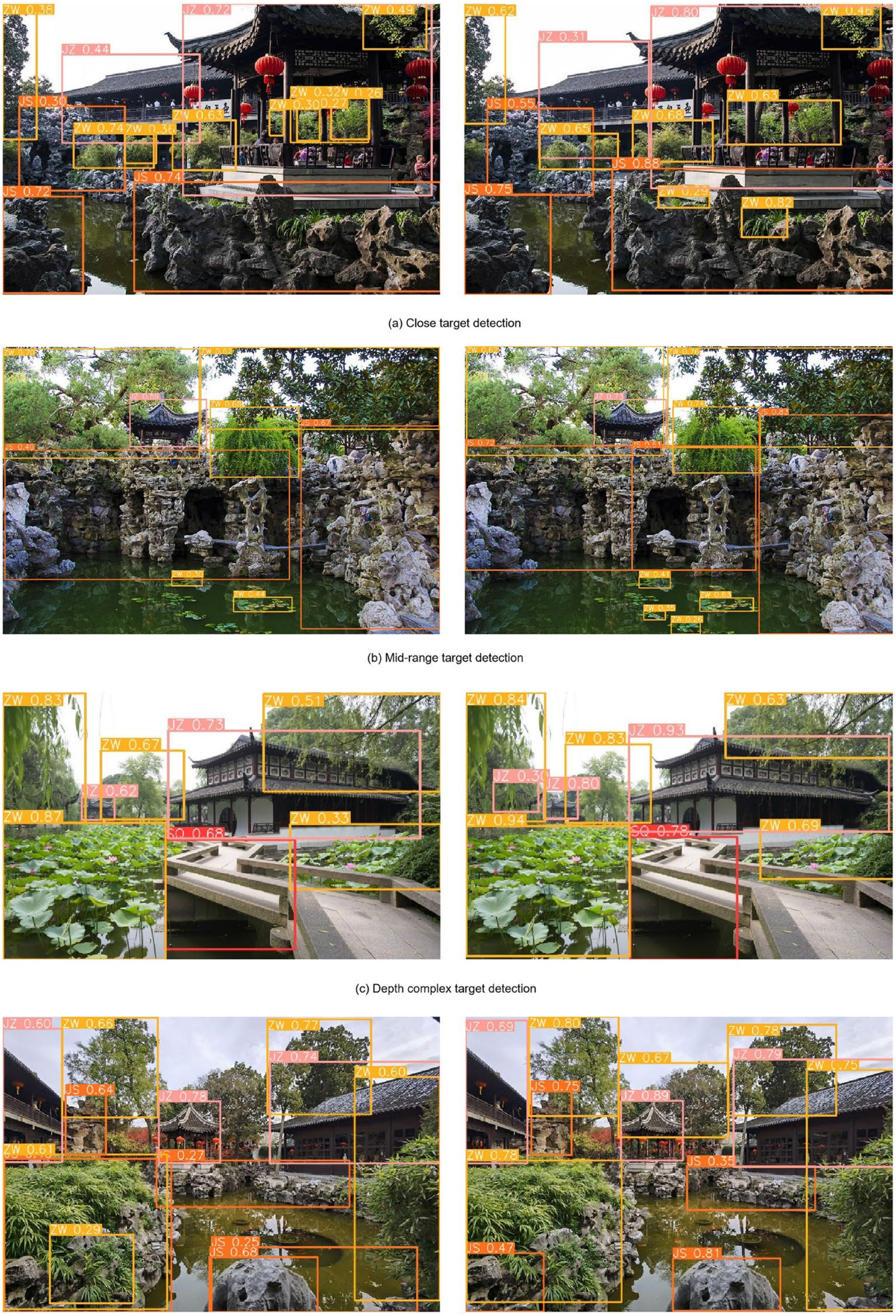
Moreover, to validate the feasibility of YOLOv8-modify, representative photographs from the test set were selected. Figure 19 illustrates the model detection comparison results of YOLOv8n-modify and YOLOv8n in various scenes. The left and right images represent the detection results of the YOLOv8n and YOLOv8n-modify models, respectively.

Figure 19a illustrates a comparison of the detection results for proximate targets between the two models. In the test image, categories JZ and JS are closely situated, with obstructions in the foreground and background. The left image omits the detection of ZW concealed within the JS, whereas the right image successfully identifies them. For the primary building scene, the confidence levels are 0.72 and 0.80 for the left and right images, respectively. For JS, the left image exhibited a confidence level of 0.32, while the right image exhibited a confidence level of 0.63.

Figure 19b shows a comparison of the detection outcomes for the mid-range targets, encompassing JZ and JS in the mid-range and ZW in the distant background. For JZ situated in the image center, the confidence levels are 0.59 and 0.73 for the left and right images, respectively. The image on the left neglects the detection of ZW in the foreground pool, whereas the image on the right identifies it.

Figure 19c juxtaposes the detection results of intricate garden landscape objects with a depth of field, wherein SQ serves as a depth-extending element and ZW partially obscures JZ. The right image surpasses the left in target recognition; it exhibits a confidence level 0.1 higher for JS recognition, 0.2 higher for JZ recognition, and 0.36 higher for ZW recognition.

Figure 19d contrasts the detection results for densely packed and complex targets, where JZ exhibits varied orientations and a pronounced depth sense, and the image comprises close, medium, and long shots. ZW was concealed in JS, JS was obscured in water, and the detection targets were multifaceted. The right image outperforms the



(a) Close target detection

(b) Mid-range target detection

(c) Depth complex target detection

(d) Complex background target detection

Fig. 19 Comparison of YOLOv8n and YOLOv8n-modify in various scenes

left in identifying diverse target types; it exhibits a JZ confidence level 0.11 higher, JS recognition of 0.56 higher, and ZW recognition 0.17 higher than those of the left image.

Model limitations and disadvantages

Limited detection of small targets: The YOLOv8n-modify model, despite adopting multilevel and multiscale feature extraction, exhibits a relatively weak detection ability for small targets, potentially omitting some minute rockeries and distant stone bridges.

Robustness to Occlusion and Complex Scenes: The performance of a model may be compromised when a target is occluded or situated in a complex background environment. For instance, stone bridges or buildings partially concealed by green trees are not well recognized, and water bodies in gardens are not accurately identified owing to reflections and transparency or confusion with the surrounding environment.

Conclusions and future work

Conclusions

(1) A specialized dataset comprising 4890 images, which includes internet-crawled pictures and on-site photographs, covering the "Four Famous Gardens" and other renowned gardens, reflecting various angles and lighting conditions, has been constructed. The dataset was annotated using rectangular boxes and category labels and manual annotation was performed using LabelImg. The dataset was randomly divided using the holdout method and the image size was adjusted and normalized. Data augmentation strategies such as image mirroring, random flipping, random graying, Gaussian noise addition, and random occlusion have been employed to enhance the model's generalization ability and robustness.

(2) By enhancing the network structure of YOLOv8, the detection accuracy was improved, thereby addressing the identification problem of complex garden scenes. The DBB module in the backbone optimizes feature extraction through a multi-branch structure and multiscale convolution, improves detection accuracy, and maintains efficient operation during the inference phase. The replacement of BiFPN in the neck enhances feature fusion and improves the retention rate of small-object information in target detection. The integration of an attention mechanism called DyHead into the head layer enhances the representational capability for object detection.

(3) The experimental results demonstrated that the improved model exhibited superior accuracy and robustness in the complex scenes of Jiangnan private gardens. The improved model demonstrated a precision of 66.1%, a recall of 46.7%, and an mAP50 of 57.1%, validating the effectiveness of the proposed method for object detection.

Future work

(1) Enhancing model accuracy and robustness: Employ strategies to improve the model's detection of small targets and its robustness in occlusions and complex scenes, such as introducing a refined feature pyramid network and exploring the use of contextual information and multiscale attention mechanisms.

(2) Combining semantic information: Target detection is integrated with the cultural significance and historical value of the scene by combining cultural heritage data and historical materials to enhance the understanding and recognition ability of garden landscape elements.

(3) Multimodal data fusion: Consider fusing other types of data, such as lidar data and thermal infrared data, with image data to improve the accuracy and robustness of target detection.

(4) Lightweight and acceleration models: Lightweight and acceleration algorithms, such as model compression, quantization, and pruning technologies, can reduce the storage and computational overhead of the model and achieve faster target detection.

In summary, the key to enhancing model accuracy and robustness lies in the integration of semantic information, multimodal data, and the development of more streamlined models. Such technological innovations can effectively propel research in target detection, thereby accelerating the conservation efforts for traditional garden heritage.

Acknowledgements
Not applicable.

Author contributions
CG: Conceptualization, Methodology, Software, Validation, Formal analysis, Data Curation, Writing – Original, Draft Visualization, Visualization. QZ: Conceptualization, Methodology, Software, Writing – Original, Visualization, Supervision. ZT: Conceptualization, Methodology, Software, Visualization. GZ: Conceptualization, Methodology, Software, Draft Visualization. SG: Conceptualization, Methodology, Software, Draft Visualization. EK and TS: Methodology, Software, Resources, Writing—Review & Editing, Supervision. All authors have read and agreed to the published version of the manuscript.

Funding
Authors would like to acknowledge the financial support of Chinese Ministry of Education Humanities and Social Sciences Research Youth Fund Project (Grant No. 23YJC760101) and Scientific Research Fund of Zhejiang Provincial Education Department (Grant No. Y202351058).

Availability of data and materials

The datasets generated during the current study are available from the corresponding author on reasonable request.

Code availability

Not applicable.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 11 November 2023 Accepted: 10 January 2024
Published online: 29 January 2024

References

1. Qu H. A brief analysis of the gardening art of Lingnan private gardens—compared with Ming and Qing Dynasties Jiangnan Private Gardens. J South China Agric Univ Soc Sci Ed. 2007;6(3):118–21.
2. Li Z, Sun J, Cao N, Li W. The extension of Jiangnan private garden gardening art in modern residential area design. J Northwest For Univ. 2013;28(3):220–3.
3. Yuan Yixin, Liu S. Analysis of the evolution Mechanism of Individual Private Gardens in Jiangnan during the Ming and Qing Dynasties Based on Dynamic Perspectives. Huazhong Architecture, 2021, 39(02): 30–93. https://doi.org/10.13942/j.cnki.hzjz.2021.02.019
4. Zhang Zhihao. Analysis of the Architectural Art of Jiangnan private gardens: Taking Hu Xueyan's Former Residence as an Example. Art Res; 2020(04): 12–13https://doi.org/10.13944/j.cnki.ysyj.2020.0226
5. Wang L. Research on the gardening art of traditional Jiangnan private gardens under the aesthetic thought of Song Dynasty landscape painting. Master's thesis, Qilu University of Technology; 2020.
6. Qi Yu, Zhang Wankun. A Comparitive Study of Garden Art of Lingnan private gardens and Jiangnan Private gardens. Fashion of Tomorrow; 2020(08): 49–50.
7. Marr D, Hildreth E. Theory of edge detection. Proc R Soc Lond Ser B Biol Sci. 1980;207(1167):187–217.
8. Lowe DG. Object recognition from local scale-invariant features. In: Proceedings of the seventh IEEE international conference on computer vision, vol. 2. IEEE; 1999. p. 1150–7.
9. Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05), vol. 1. IEEE; 2005. p. 886–93.
10. LeCun Y, Kavukcuoglu K, Farabet C. Convolutional networks and applications in vision. In: Proceedings of 2010 IEEE international symposium on circuits and systems. IEEE; 2010. p. 253–6.
11. Modarres C, Astorga N, Droguett EL, Meruane V. Convolutional neural networks for automated damage recognition and damage type identification. Struct Control Health Monit. 2018;25:e2230. https://doi.org/10.1002/stc.2230.
12. Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2014. p. 580–7.
13. Ren S, He K, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. Adv Neural Inf Process Syst; 2015, 28.
14. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, real-time object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. p. 779–88.
15. Terven J, Cordova-Esparza D. A comprehensive review of YOLO: from YOLOv1 to YOLOv8 and beyond. arXiv preprint arXiv:2304.00501. 2023.
16. Fang Y, Liao B, Wang X, Fang J, Qi J, Wu R, Niu J, Liu W. You only look at one sequence: rethinking transformer in vision through object detection. Adv Neural Inf Process Syst. 2021;34:26183–97.
17. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, et al. An image is worth 16x16 words: transformers for image recognition at scale. arXiv preprint arXiv:2010.11929. 2020.
18. Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A, Zagoruyko S. End-to-end object detection with transformers. In: European conference on computer vision. Springer; 2020. p. 213–29.
19. Zhang Z, Lu X, Cao G, Yang Y, Jiao L, Liu F. Vit-yolo: transformer-based YOLO for object detection. In: Proceedings of the IEEE/CVF international conference on computer vision; 2021. p. 2799–808.
20. Soeb MJA, Jubayer MF, Tarin TA, Al Mamun MR, Ruhad FM, Parven A, et al. Tea leaf disease detection and identification based on YOLOv7 (YOLO-T). Sci Rep. 2023;13(1):6078.
21. Duan C, Luo S. Design of pedestrian detection system based on OpenCV. In: 2022 4th international conference on artificial intelligence and advanced manufacturing (AIAM). IEEE; 2022. p. 256–9.
22. Zhang X, Feng Y, Zhang S, Wang N, Mei S. Finding nonrigid tiny person with densely cropped and local attention object detector networks in low-altitude aerial images. IEEE J Sel Top Appl Earth Observ Remote Sens. 2022;15:4371–85.
23. Jiang C, Ren H, Ye X, Zhu J, Zeng H, Nan Y, et al. Object detection from UAV thermal infrared images and videos using YOLO models. Int J Appl Earth Obs Geoinf. 2022;112:102912.
24. Tceluiko DS. Garden space. Morphotypes of private gardens of Jiangnan region. IOP Conf Ser Mater Sci Eng. 2020;775(1):012058.
25. Zheng J. Art and the shift in garden culture in the Jiangnan Area in China (16th–17th Century). Asian Cult Hist. 2013;5(2):1.
26. Wang C. Research on gardening art from the perspective of different aesthetic forms—taking the example of private gardens in Jiangnan of the Ming Dynasty. Highlights Art Des. 2023;3(2):104–9.
27. Reis D, Kupec J, Hong J, Daoudi A. Real-time flying object detection with YOLOv8. arXiv preprint arXiv:2305.09972. 2023
28. Zou MY, Yu JJ, Lv Y, Lu B, Chi WZ, Sun LN. A novel day-to-night obstacle detection method for excavators based on image enhancement and multi-sensor fusion. IEEE Sens J. 2023;23:10825–35.
29. Wang N, Liu H, Li Y, Zhou W, Ding M. Segmentation and phenotype calculation of rapeseed pods based on YOLO v8 and mask R-convolution neural networks. Plants. 2023;12(18):3328.
30. Bochkovskiy A, Wang CY, Liao HYM. YOLOv4: optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934. 2020.
31. Redmon J, Farhadi A. YOLO9000: better, faster, stronger. In Proceedings of the 2017 IEEE conference on computer vision and pattern recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; 2017. p. 6517–25.
32. Zhu X, Lyu S, Wang X, Zhao Q. TPH-YOLOv5: improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios. In: Proceedings of the IEEE/CVF international conference on computer vision, Montreal, BC, Canada, 11–17 October 2021; 2021.
33. Lou H, Duan X, Guo J, Liu H, Gu J, Bi L, Chen H. DC-YOLOv8: small-size object detection algorithm based on camera sensor. Electronics. 2023;12(10):2323.

Publisher's Note